理學博士學位論文

*Streptomyces coelicolor*에서 전사 조절 인자인 BldD의 carboxy 말단 지역 단백질의 기능 및 구조 분석

Functional and structural analyses of the carboxy-terminal domain of the transcription factor BldD in *Streptomyces coelicolor* A3(2)

2014年 2月

서울大學校 大學院

生命科學部

金 正 穆

*Streptomyces coelicolor*에서 전사 조절 인자인 **BldD**의 **carboxy** 말단 지역 단백질의 기능 및 구조 분석

指導敎授　姜 思 旭

이 論文을 理學博士學位論文으로 提出함

2013年　11月

서울大學校

生命科學部

金 正 穆

金正穆의 理學博士學位論文을 認准함

2013年 12月

委 員 長 _____

副委員長 _____

委　　員 _____

委　　員 _____

委　　員 _____

Functional and structural analyses of the carboxy-terminal domain of the transcription factor BldD in *Streptomyces coelicolor* A3(2)

by

Jeong-Mok Kim

Advisor:

Professor Sa-Ouk Kang, Ph. D.

A Thesis Submitted in Partial Fulfillment

of the Requirements for

the Degree of Doctor of Philosophy

February, 2014

School of Biological Sciences

Seoul National University

ABSTRACT

BldD is a DNA-binding protein with 167 amino acids and acts as a repressor for key developmental genes in *Streptomyces coelicolor*. Although extensive researches have emphasized the importance of BldD in developmental processes of *Streptomyces*, distinct regulatory mechanism of BldD has not been well understood yet. The N-terminal domain of BldD (residues 1-79, BldD-NTD) has clear functions that mediate DNA-binding and dimerization, but the function has not been defined for the C-terminal domain of BldD (residues 80-167, BldD-CTD). Therefore, the function of BldD-CTD could more likely be related with the regulatory mechanism of BldD. In this study, backbone and side-chain NMR assignments of the recombinant BldD-CTD protein could be achieved by a series of NMR experiments on a [^{13}C, ^{15}N]-enriched protein sample. The secondary structure prediction by CSI and TALOS+ analysis using the assigned chemical shift data identified that the BldD-CTD adopts a βααβααα fold. From backbone and side-chain assignments of the recombinant BldD-CTD, NOE cross-peaks assignments were also completed for 3D-structure calculation. The determined solution structure of BldD-CTD is very similar to winged-helix domains in spite of different topology. But, DNA-binding of BldD-CTD is not structurally

favorable because of slightly negative-charged surface and additional helical region. As removal of additional helical region did not show any functional difference to native protein, assessed by gel mobility shift assays and *in vivo* complementation experiments, it is the anionic property of the BldD-CTD that appears to be mainly responsible for its inability to bind DNA. Conserved surface analysis of BldD-CTD revealed that highly conserved hydrophobic patch surrounded by charged residues is located opposite to helix-turn-helix region. These structural features suggest that BldD-CTD constitutes a novel fold of winged-helix domain involved in protein-protein interaction and this interaction could be directly related to the regulatory mechanism of BldD.

Key words: *Streptomyces coelicolor*, Differentiation, BldD, DNA-binding protein, NMR, Solution structure, Winged-helix domain, Protein-protein interaction

CONTENTS

LIST OF TABLES

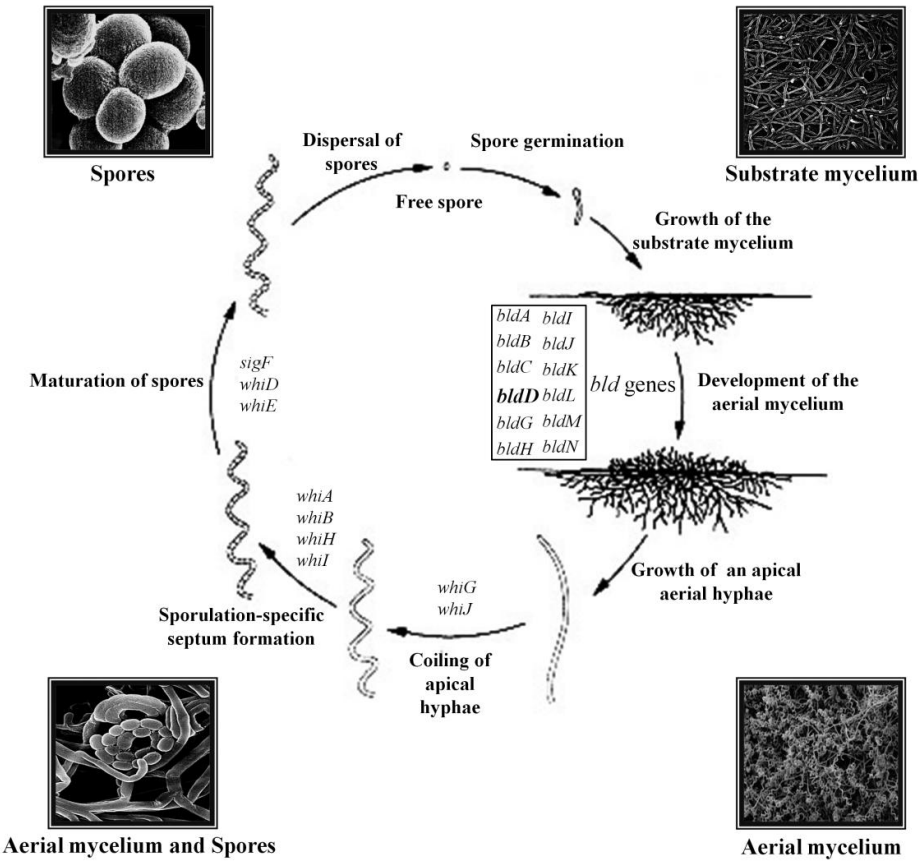
LIST OF FIGURES

I. INTRODUCTION

1. Morphological differentiation of *Streptomyces coelicolor*

Streptomyces coelicolor is spore-forming, filamentous soil bacterium. Spore formation is accomplished through a complex process of morphological differentiation. Under favorable condition, one or two germ tubes emerge from a spore and grow by tip extension and branch formation to give rise to a substrate mycelium. In response to nutrient depletion, aerial hyphae, which grow out into the air constituting a fuzzy layer on the colony surface, are formed and then further differentiate to form chains of unigenomic spores (Kelemen & Buttner, 1998) (Scheme. I). Together with these morphological changes, secondary metabolites such as antibiotics and pigments are also produced (Chater, 1993).

Numerous studies revealed that two classes of genes (*whi* genes and *bld* genes) are essential for the morphological development. The *bld* mutants fail to form aerial mycelium and have a shiny ‘bald’ appearance (Merrick, 1976; Willey *et al.*, 1993). The *whi* mutants are capable of formation of aerial hyphae but lack of the grey spore pigment, of which phenotype yield ‘white’ (Chater, 1972). These mutants also show pleiotropic effects, including defects in cell-



Scheme 1. Life cycle of *Streptomyces coelicolor*. Under favorable condition, one or two germ tubes emerge from a spore and grow by tip extension and branch formation to give rise to a substrate mycelium. In response to nutrient depletion, aerial hyphae, which grow out into the air constituting a fuzzy layer on the colony surface, are formed and then further differentiate to form chains of unigenomic spores The *bld* genes are required for the erection of aerial hyphae and the *whi* genes are required for sporulation (modified from Kieser *et al.*, 2000).

cell signaling and carbon catabolite repression (Kelemen & Buttner, 1998; Merrick, 1976; Willey *et al.*, 1993).

1.1. *whi* mutants

Sporulation-deficient mutants of *S. coelicolor* which showed 'white' phenotype have been mapped genetically, and it has been revealed that all of these mutants have one of the eight separate loci: *whiA*, *whiB*, *whiD*, *whiE*, *whiG*, *whiH*, *whiI*, and *whiJ* (Chater, 1972; Chater & Merrick, 1976; Chater *et al.*, 1989; Davis & Chater, 1992). The *whiA* locus is apparently part of a complex operon and encodes a previously unknown kind of protein with the eukaryotic glutathione peroxidase-related motifs (Ainsa *et al.*, 2000). And *whiB* encodes an unusually small protein (~ 9.9 kDa), the C-terminus of which fulfills some criteria for DNA binding domain that has been suggested for eukaryotic transcription factors (Chater, 1993; Davis & Chater, 1992). Recent growth in the mycobacterial DNA sequence database has led to the discovery of a family of mycobacterial genes encoding WhiB-like proteins. Although the biochemical roles of these homologues are unknown, primary sequence alignment shows that all of them have four conserved cysteinyl residues (CXXC motif), encouraging speculation that they may bind to metals. Perhaps the influence of WhiB on transcription of putative target genes may be responsive to changes in

redox potential (Chater, 1993). WhiD is 112 amino acids long and is a member of a family of proteins that includes WhiB (Molle *et al.*, 2000). The *whiD* mutant formed sporulation septa but failed to go on to produce mature spores. The *whiD* mutant formed spores at wild type abundance but these were unpigmented, were thin-walled, and showed frequent lysis. The *whiE* genes encode proteins that closely resemble the components of type II polyketide synthases, which are involved in the synthesis of a variety of aromatic antibiotics, including tetracenomycin from *Streptomyces glaucescens*, granaticin from *Streptomyces violaceoruber*, oxytetracycline from *Streptomyces rimosus*, and actinorhodin from *S. coelicolor* itself (Kelemen *et al.*, 1998). The sigma factor encoded by *whiG* belongs to flagellar group of sigma factors (Chater *et al.*, 1989; Tan *et al.*, 1998). Since other flagellar group sigma factors such as *B. subtilis* σ^{D} and *E. coli* σ^{F} (*FliA*) are regulated by anti-sigma factor, σ^{whiG} is possibly controlled also by an anti-sigma factor that might be responsive to an event associated with the emergence of aerial hyphae (Losick & Shapiro, 1993). The observed transcription of *whiG* in liquid cultures and early surface cultures devoid of aerial hyphae implies that σ^{whiG} sigma factor is itself present before formation of aerial hyphae (Kelemen *et al.*, 1996). The WhiH protein resembles members of GntR family (FadR, LatR and RdhR) repressors (Ryding *et al.*, 1998). Many members of this family respond to

organic acids that occur in intermediate metabolism. This suggests that WhiH may sense one of these molecules whose concentrations fluctuate significantly in different metabolic states. *In vitro* and *in vivo* evidences strongly indicate that the *whiH* promoter is transcribed by the σ^{whiG}-containing RNA polymerase (Ryding *et al.*, 1998). The onset of *whiH* transcription also coincides with perceptible aerial growth and is much later than that of *whiG* transcription (Ryding *et al.*, 1998). DNA that complements several *whiI* (closely linked to *cysD*) mutants encodes a member of the response regulator family of proteins (Ryding *et al.*, 1999). This gene has putative promoter sequences similar to those of the σ^{whiG}-dependent promoters. The deduced protein of this gene has all the conserved features needed; the adjacent gene appears to encode a membrane-located enzyme of this kind. If this is indeed the *whiI* locus, the membrane-associated kinase domain may permit it to detect signals in or outside of the membrane.

1.2. *bld* genes

At least 10 *bld* loci, *bldA, bldB, bldC, bldD, bldF, bldG, bldH, bldI, bldK, bldL,* have been found in *S. coelicolor* (Champness, 1988; Chater & Merrick, 1976; Merrick, 1976; Nodwell *et al.*, 1996; Willey *et al.*, 1993; Willey *et al.*, 1991). The highly pleiotropic phenotype of these mutations suggests that

these genes are involved with an early stage on the initiation of development. Interestingly, when certain pairs of *bld* mutants are grown on rich media in close proximity, one mutant can trigger the aerial hyphae formation of the other mutant (Molle & Buttner, 2000; Nodwell *et al.*, 1999; Tillotson *et al.*, 1998; Willey *et al.*, 1993; Willey *et al.*, 1991). As extracellular complementation among *bld* mutants is always unidirectional, this can be explained by the hierarchical cascade of intercellular signals for the formation of aerial hyphae. That is, one mutant that is higher in the hierarchy and acts as a donor can give the signals to the other mutant acting as a recipient through unidentified signaling mechanism. Such analyses have suggested the following hierarchy of extracellular complementation groups:

[bldJ] < [bldK – bldL] < [bldA – bldH] < [bldG] < [bldC] < [bldD – bldM]

All blocked mutants to the left are complemented by those to the right, and those in the same group do not interact with each other and display the same pattern of complementation. These mutants are known to be unable to produce a small hydrophobic molecule, SapB that contributes to the erection of the aerial hyphae by reducing the surface tension at the colony surface (Tillotson *et al.*, 1998; Willey *et al.*, 1991). Therefore, in the extracellular complementation of *bld* mutants, one donor mutant renders the recipient mutant to restore the ability to produce the final product, SapB.

Only two *bld* genes, *bldA* and *bldK*, have been characterized at the molecular level. The *bldA* alleles reside in a gene for a leucyl-tRNA that recognized the UUA codon (Lawlor *et al.*, 1987). UUA codon is a rare codon in *Streptomyces*, and it has been suggested that this tRNA is involved in the translation of regulatory genes involved in antibiotic production and morphogenesis (Fernandez-Moreno *et al.*, 1991; Leskiw *et al.*, 1991). It has been shown that the *bldK* locus consists of five adjacent ORFs that specify homologues of the subunits of the oligopeptide permease family of ATP-binding cassette (ABC) membrane-spanning transporters (Nodwell *et al.*, 1996). One of the most intriguing but poorly understood aspects of the *bld* phenotype is that growth on poor carbon sources is sufficient to restore partially the morphological and antibiotic defects of most of these mutants. Both genes for *bldB* and *bldD* were isolated and characterized as small proteins with helix-turn-helix signature of LysR family regulatory proteins (Elliot *et al.*, 1998; Pope *et al.*, 1998), but the actual roles in aerial mycelium formation had remained unknown.

2. General description of *bldD*

Among the *bld* mutants, the *bldD* mutant has mostly severe pleiotropic defects (Merrick, 1976). On minimal medium containing glucose as carbon

source, the *bldD53* mutant has a fragmented surface lacking aerial structures on its colonies and produces none of the four well-characterized antibiotics known to be produced by *S. coelicolor*. The morphological defect of mutant is the case for many of the *bld* mutants, but the loss of antibiotic production is not. Its defect is overcome by growth on minimal medium containing mannitol as the carbon source.

The *bldD* gene product, BldD, consists of 167 amino acids with a calculated molecular weight of 18,167 Da and the identified *bldD* mutant has a point mutation at position 62 from Tyr to Cys (Elliot *et al.*, 1998). BldD binds to its own promoter (Elliot & Leskiw, 1999) and regulatory regions of key developmental genes (Elliot *et al.*, 2001) including *bldN*, which encodes a sigma factor involved in aerial hyphae formation (Bibb *et al.*, 2000), and *whiG*, which encodes a sigma factor related in spore formation (Chater *et al.*, 1989). The stress-responsive genes, *sigH*, is also regulated by BldD, which indicates that BldD is a key regulator connecting stress to developmental processes (Kelemen *et al.*, 2001). The DNaseI footprinting analyses for some of the BldD target genes suggested the imperfect inverted repeat, AGTgA $(n)_m$ TCACc, as the consensus sequence for BldD binding (Elliot *et al.*, 2001). The locations of these binding sites and the differences of transcriptional levels for the putative BldD target genes between wild type and *bldD* mutants revealed that BldD acts

as a repressor during vegetative growth (Elliot & Leskiw, 1999; Elliot *et al.*, 2001). Recently, chromatin immunoprecipitation-microarray analysis (ChIP-chip) has further extended the range of BldD-regulated genes, strengthening the importance of BldD in the life cycle of *Streptomyces coelicolor* (den Hengst *et al*., 2010). But, how BldD regulates various target genes in response to developmental signal is still unknown.

3. Aims of this study

BldD exists predominantly as a homodimer in solution (Elliot *et al*., 2003) and each subunit is composed of two structurally independent domains (Lee *et al*., 2007). The N-terminal domain (residues 1-79, BldD-NTD) is clearly responsible for DNA-binding through its helix-turn-helix motif belonging to XRE (Xenobiotic Response Element) family (Kim *et al*., 2006; Lee *et al*., 2007). Contrarily, the function of the C-terminal domain (residue 80-167, BldD-CTD) has not been understood yet, except that it has a helical structure and acts as a monomer in solution (Lee *et al*., 2007). In the intact BldD dimer, no significant interdomain interaction has been observed between the two distinct domains. Furthermore, isolated BldD-NTD showed no functional defect in dimerization and DNA-binding (Lee *et al*., 2007). Many times of attempts to develop the single crystal from the full-length BldD has ended in failure due to degradation

of BldD-CTD. Thus, the structural information of BldD-CTD acquired by NMR in solution is essential to understand the precise regulatory action of BldD directly linked to the developmental processes in *Streptomyces coelicolor*. In this work, almost complete NMR assignments for BldD-CTD were accomplished and three dimensional BldD-CTD structure with high quality was determined. The BldD-CTD structure adopts a novel fold similar to winged-helix domain, but was not compatible with DNA-binding. Therefore, BldD-CTD structure would provide new insight for regulatory mechanism of BldD, and extensively to role of winged-helix domain.

II. Materials and Methods

1. Bacterial strains, media and culture conditions

S. coelicolor A3(2) strain was grown and maintained as described by Hopwood *et al*. (1985). For liquid culture, pre-germinated spores (about 10^8~10^9 spores/100ml broth) were grown in YEME liquid medium (1% glucose, 0.5% Bacto-peptone, 0.3% malt extract (Difco), 0.3% yeast extract (Difco)) containing 34% sucrose at 30°C with vigorous shaking. Pre-germinated spores (about 10^7) or patches of mycelia were inoculated on R2YE (10.3% sucrose, 1% glucose, 1.01% $MgCl_2$, 0.024% K_2SO_4, 0.001% casamino acid (Difco), 0.5% yeast extract, 1.43% (~20 mM) TES (N-tris[hydroxymethyl]methyl-2-aminoenthanesulfonic acid, pH 7.0), 20 mM $CaCl_2$, 0.005% K_2HPO_4, and 0.3% proline) for analysis of phenotype or NA (0.8% nutrient broth (Difco)) agar media for selection of mutant strains or MS (2% D-mannitol, 2% soybean flour and 2% agar) for spore preparation and grown at 30°C. For preparation of mycelia on solid media, 10^7 pre-germinated spores were inoculated on R2YE over laid with cellophane discs. The morphological developmental process was confirmed routinely by microscopic observations. All the recombinant DNAs were introduced into competent *E. coli* DH5α (Sambrook *et al*., 1989). For

conjugation of DNA into *S. coelicolor*, methylation-deficient *E. coli* ET12567 containing pUZ8002 which is a compatible *oriT*-containing plasmid. For overexpression of recombinant proteins using T7 polymerase-based system, *E. coli* BL21(DE3)pLysS (Novagen) was used. *E. coli* cells were grown at 37°C in LB (1% tryptone (Difco), 0.5% yeast extract, and 1% NaCl) or M9 (1.79% $Na_2HPO_4 \cdot 12H_2O$, 0.3% KH_2PO_4, 0.1% NH_4Cl, 0.1 mM $CaCl_2$, 0.2 M $MgSO_4$, 1 μg·mL^{-1} Thiamine, 0.3% Glucose) supplemented with appropriate antibiotics.

2. Construction of plasmids for overexpression of BldD, BldD-CTD, and their truncated forms

Oligonucleotide primers were designed from the sequence of *bldD* coding region with restriction sites including NdeI and BamHI. DNA were amplified by PCR (Polymerase Chain Reaction) using *S. coelicolor* A3(2) chromosomal DNA as a template. PCR reactions were performed in a Perkin-Elmer Cetus thermal cycler for 30 cycles with the following optimized conditions: denaturation at 98°C for 20 seconds, annealing at 68°C for 20 seconds, extension at 72°C for 1 min. Amplified DNA were separated by 1% agarose gel and visualized on UV-illuminator. The DNA band matching to expected size was sliced, extracted and cleaned by gene cleaning kit (MoBio).

Purified DNAs were introduced into pGEM-Teasy vector (Promega) and were sequenced for the verification of amplified sequences. The pGEM-Teasy vector containing verified gene coding region corresponding to BldD or BldD-CTD was excised by restriction enzymes, NdeI and BamHI. The excised DNA was purified and cleaned with same procedures for PCR products. The pET-15b vector (Novagen) was also excised by same restriction enzymes, purified, and cleaned with same procedures. The resultant excised pET-15b vector and DNA corresponding to coding region for BldD or BldD-CTD were ligated by T4 ligase (Fermentas) following provided protocol. The plasmids for the truncated forms of BldD or BldD-CTD were also constructed following same procedures for construction of plasmid containing coding region of BldD or BldD-CTD.

3. *In vivo* complementation experiments

To construct a plasmid for complementation experiment, the DNAs containing promoter region of *bldD* and ORF for BldD and its fragments were amplified with PCR and cloned into pGEM-Teasy vector. DNA fragments from digestion of EcoRI were ligated to into pSET162, which has an insertion of a thiostrepton resistance marker at SphI site of pSET152 (Bierman *et al*., 1992). The constructed pSET162 vectors for *in vivo* complementation were verified by sequencing the purified vectors. The pSET162 derivatives were introduced into

methylation-deficient ET12567 containing the pUZ8002 vector for conjugation. The *bldD* deletion mutant was conjugated with the transformed cells derived from ET12567/PUZ8002. The integration of recombinant construct to *bldD* deletion mutant was verified by resistance for apramycin and thiostrepton on NA solid medium and PCR analysis of genomic DNA from the pSET162-integrated candidates. For phenotypic analysis, wild type and mutant strains were grown on R2YE solid medium at 30°C and monitored after 5 days.

4. Overexpression and purification of BldD, BldD-CTD, and their truncated forms

A freshly prepared *E. coli* BL21(DE3)pLysS was transformed with pET-15b or pET-3a vector containing sequence corresponding to BldD or BldD-CTD or its fragments. The transformed cells were grown in LB medium containing 50 μg/ml ampicillin and 30 μg/ml chloramphenicol at 37°C. When the cell density reached an OD_{600} of about 0.5, protein expression was induced with 1 mM IPTG and cells were incubated at 22°C for 15 h. The cells were harvested by centrifugation, resuspended in 100 ml binding buffer containing 1 mM PMSF and disrupted by sonication. After centrifugation for removing debris of cells and insoluble fraction, the supernatant was applied to a nickel affinity column. The column was washed with 5 bed volume of binding buffer

(5 mM imidazole, 100 mM NaCl, and 10 mM Tris-HCl, at pH 7.9) and sequentially with 10 bed volume of washing buffer (25 mM imidazole, 100 mM NaCl, and 10 mM Tris-HCl, at pH 7.9). The proteins were eluted with 3 bed volume of elution buffer (500 mM imidazole, 100 mM NaCl, and 10 mM Tris-HCl, at pH 7.9). The eluent from Ni-affinity column was concentrated by ultrafiltration using YM-1 membrane (Millipore). The concentrated protein was applied to PD-10 column (GE healthcare) for exchanging buffer to thrombin cleavage buffer. The eluent in thrombin cleavage buffer (20 mM Tris-HCl buffer containing 150 mM NaCl and 2.5 mM $CaCl_2$, at pH 8.4) was mixed with 2 unit of thrombin (Novagen) and incubated at 22°C for 18 h to cleave the N-terminally tagged histidines, followed by the removal of thrombin and other impurities via the sequential application of nickel-affinity. For further purification of proteins and buffer-exchange, eluent was applied to gel-filtration chromatography (HiLoad 16/60 Superdex^{TM} 75, GE healthcare) and concentrated in 20 mM Tris-HCl buffer (pH 7.8) containing 150 mM NaCl. Purified proteins were qualified with tris-tricine polyacrylamide gel electrophoresis and quantified with Protein assay kits (BioRad) based on the method by Lowry.

5. Immunoblot analysis of *S. coelicolor*

The collected mycelia were resuspended in 20 mM Tris-HCl buffer (pH 7.8) containing 150 mM NaCl, 2 mM EDTA, 1 mM PMSF and 1% (v/v) Triton X-100. The resuspended mycelia were disrupted by sonication and clarified by centrifugation at 4°C. The concentration of the total protein in the crude cell extract or purified proteins was accurately quantified using Protein assay kits (BioRad) based on the method by Lowry. The 40 μg of crude extract was applied on 15.4% Tris-Tricine polyacrylamide gel electrophoresis with constant current (30 mA). The polyacrylamide gel was transferred to PVDF (Millipore) membrane using TransBlot system (BioRad) at 150 mA for 90 min. The transferred membrane was washed and blocked with three times of 10 ml Tris-buffered saline (TBS; 10 mM Tris-HCl, at pH 7.5, and 150 mM NaCl) containing 0.05% Tween20 (TBST) with 1% skim milk (Difco). Membranes were incubated with anti-BldD-NTD antiserum (from rabbit) or anti-FLAG-M2 monoclonal antibody (sigma) in TBST containing 1% skim milk for 12 hours (ratio 1:2000 for anti-BldD-NTD and 1:10000 for anti-FLAG-M2). Non-specific bound or unbound antibodies were removed by repeated washing (at least three times) with TBST. The membrane was incubated for 60 min with secondary antibody solution (anti-rabbit or anti-mouse IgG antibody-alkaline phosphatase conjugate, diluted to 1:10000 in TBST). The membrane was then washed for 15 min with three changes of TBST and rinsed with alkaline

phosphatase buffer (100 mM Tris-HCl, at pH 9.5, 100 mM NaCl, and 10 mM $MgCl_2$). The membrane was visualized through incubation in 10 ml of alkaline phosphatase buffer containing 8 mg of BCIP and 16 mg of NBT at RT.

6. Gel mobility shift assays

To prepare the probes for gel shift assay, complementary oligonucleotides were synthesized as follows:

p*whiG*I-F 5′(6-FAM) AAGGTGTTCGAGTGATCACCCAGAGCGA 3′

p*whiG*I-R 5′ ATCGCTCTGGGTGATCACTCGAACACCT 3′

p*whiG*II-F 5′(6-FAM) AAGTCCAGTCACGCTACGCTCACGATGA 3′

p*whiG*II-R 5′ ATCATCGTGAGCGTAGCGTGACTGGACT 3′

p*bdtA*-F 5′(6-FAM) AGCACGCAGCGACGAAGAGTCACCGGAA 3′

p*bdtA*-R 5′ ATTCCGGTGACTCTTCGTCGCTGCGTGC 3′

p*bldN*I-F 5′(6-FAM) ACAGTGCCTGCACGAAGCGTTATTCTCCT 3′

p*bldN*I-R 5′ AAGGAGAATAACGCTTCGTGCAGGCACTG 3′

p*bldN*II-F 5′(6-FAM) ACGGGTGAATGGTTCCGTACTGCACGTG 3′

p*bldN*II-R 5′ ACACGTGCAGTACGGAACCATTCACCCG 3′

p*bldD*-F 5′(6-FAM) AAGCAGAGTAACGCTGCGTAACCTCACA 3′

p*bldD*-R 5′(6-FAM) ATGTGAGGTTACGCAGCGTTACTCTGCT 3′

Each oligonucleotide of forward direction (annotated as F) was labeled with 6-FAM to its 5′ end. These oligonucleotides were annealed in an annealing buffer containing 20 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 50 mM NaCl. The annealing mixture was heated to 100°C for 5 min and then slowly cooled over 80 min to 30°C. These annealed oligonucleotides were precipitated with 1/10 volume of $LiCl_2$ and 2 volume of absolute ethanol, followed by centrifugation and washing pellets with 70% ethanol. The annealed probes were qualified through 8% native PAGE and quantified with absorbance at 260 nm.

Gel mobility shift assays were performed using 6-FAM-labeled probes and purified proteins. The BldD, BldD-CTD, and their truncated forms were incubated with 6-FAM-labeled probe at 30°C for 20 min in binding buffer consisting of 10 mM Tris-HCl (pH 7.8), 150 mM NaCl, 2 mM dithiothreitol, 1 μg of poly (dI-dC), and 10% glycerol. The DNA-protein complexes were run out on an 8% native polyacrylamide gel for 1 hrs at 20 mA. The gel was visualized with 60s exposure time in SYBR mode using LAS3000 (Fujifilm).

7. NMR spectroscopy

7.1 Stable isotope labeling

For uniformly labeling of BldD-CTD (residues 80–167) with stable

isotope, *E. coli* BL21(DE3)pLysS cells containing pET-15b vector for expression of BldD-CTD were grown at 37 °C in M9 minimal media supplemented with $^{15}NH_4Cl$ and ^{13}C-glucose as the sole nitrogen and carbon sources, respectively. The same procedures for expression and purification of proteins were performed as mentioned above in section 4. Finally, 2 mM of uniformly [^{13}C,^{15}N]-enriched BldD-CTD, dissolved in a 40 mM sodium phosphate buffer at pH 6.8 and containing 93% H_2O/7% D_2O, 150 mM NaCl, and 0.05% (w/v) NaN_3, was prepared for NMR.

7.2 NMR measurement and chemical shift assignment

NMR experiments were performed at 295 K on a Bruker Biospin Avance 500 NMR spectrometer equipped with a cryoprobe at KBSI (Korea Basic Science Institute, Chungbuk, Korea). 2D-[^{1}H,^{15}N]HSQC, 2D-[^{1}H,^{13}C]HSQC, 3D-HNCACB, 3D-HN(CO)CACB, 3D-HNCO, 3D-^{15}N-TOCSY-HSQC, 3D-HCCH-COSY, and 3D-HCCH-TOCSY spectra were measured for backbone and side-chain resonance assignments. For referencing the chemical shifts, resonance frequency of DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) was measured. Chemical shifts of ^{1}H atoms were referenced directly, and those of ^{15}N and ^{13}C atoms were referenced indirectly, using the chemical shift ratio values suggested in the BMRB (Biological Magnetic

Resonance Bank: http://www.bmrb.wisc.edu).

All NMR spectra were processed using the NMRPipe (Delagio *et al*., 1995) software package and analyzed with NMRView (Johnson, 2004) program. In processing the raw FIDs (Free Induction Decay), to minimize water signal, the solvent suppression filter was applied to ^{1}H dimension of all data sets (Marion *et al*., 1989). Two-fold zero-filling was also applied to each dimension for fine resolution of data. 3D-[^{15}N]-edited NOESY and 3D-[^{13}C]-edited NOESY spectra were assigned by manual and automated assignments using CYANA 2.1 (Herrmann *et al*., 2002) program.

7.3 Secondary structure prediction and 3D-strucrue calculation

Secondary structure elements was predicted by CSI (Chemical Shift Index) (Wishart et al., 1992; Wishart et al., 1994) and TALOS+ analysis (Shen et al., 2009) with the assigned $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$ and $^{1}H^{\alpha}$ chemical shifts. Dihedral angle restraints derived from the prediction of TALOS+ and distance restraints from manual assignments were used for initial calculation of structure. Initial structures of the BldD-CTD from manual assignment were calculated only by torsion angle dynamics algorithm DYANA (Herrmann *et al*., 2002) included in CYANA program. For automated assignment of NOESY signal, CANDID module was used with initial input of manual assignments and dihedral angle

restraints. The NOESY cross-peaks and its distance restraints from automated assignment were validated by thoroughly examination of visualized each NOESY spectrum. From repetitive validation of automated assignment, almost all of the noise peaks and wrong-assigned peaks are removed or corrected. The 20 structures with the lowest target function values from the final cycle of CYANA calculation were further refined with explicit water (Linge *et al*., 2003) using CNS (Brünger *et al*., 1998) with RECOORD script (Nederveen *et al*., 2005).

7.4 Validation and deposition of calculated structures

The final 20 structures out of total 100 structures calculated were validated using the PSVS 1.5 software package (Bhattacharya *et al*., 2007). Atomic coordinates for the 20 structures with lowest energies and all assigned chemical shift values were deposited in the Protein Data Bank (PDB ID 2MC4) and Biological Magnetic Resonance Bank (BMRB accession number 19427), respectively.

7.5 Heteronuclear NOE measurement

The steady-state [^{1}H,^{15}N]-heteronuclear NOE values were acquired from the [^{1}H,^{15}N]-HSQC spectra recorded with (NOE experiment) and without

(NONOE experiment) using a proton presaturation period of 3 sec (Farrow *et al*., 1994). The NOE values of individual residues were determined from the ratio of the peak intensities with and without saturation.

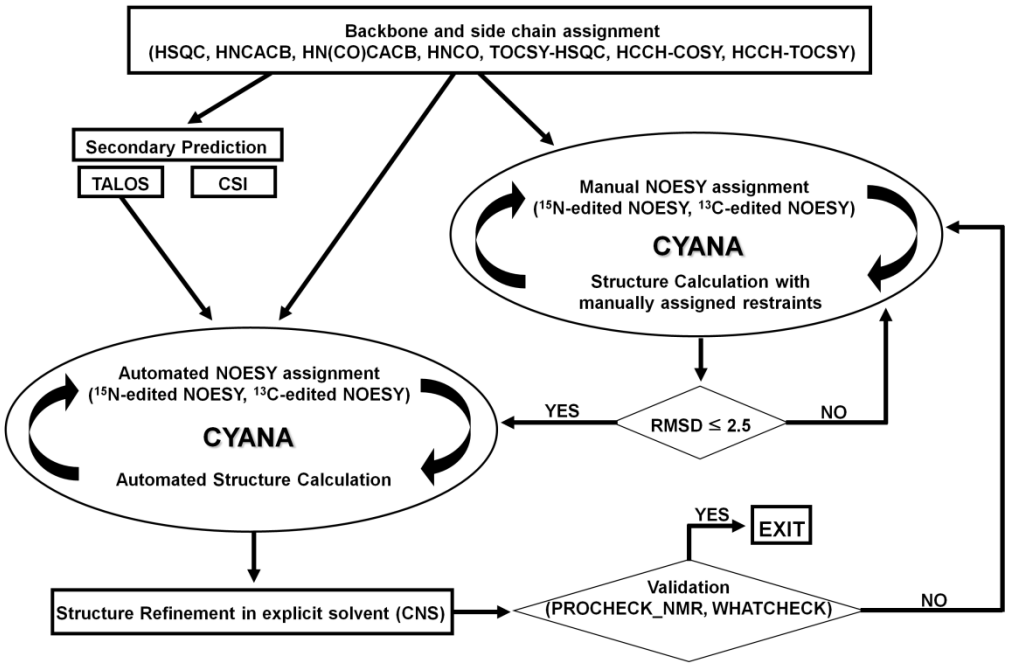


Fig. 1. Overall scheme of solution structure calculation. Based on backbone and side-chain assignment, NOE cross-peaks are assigned manually or automatically. If structure generated from manual assignment has a RMSD value below 2.5 Å, distance restraints of manual assignment can be included in automated structure calculation process of CYANA. Dihedral angle restraints from TALOS are also used in CYANA as NOE-independent restraints. To bring protein structure close to physical reality, structure generated from CYANA goes through refinement process. Final structure is validated to evaluate structure quality.

III. Data and Results

The solution 3D-structure calculation from NOESY cross-peaks has several processes summarized in Fig. 1 (Güntert, 2004). The assignment of chemical shifts using spectra from NMR measurements of proteins is the first step for analysis of NOESY cross-peaks.

1. Verification of peak clusters for backbone assignments

2D-[^{1}H/^{15}N]HSQC, 3D-^{15}N-TOCSY-HSQC and a series of triple resonance spectra including 3D-HNCACB, 3D-HN(CO)CACB, 3D-HNCO were measured on uniformly [^{13}C/^{15}N]-labeled BldD-CTD. As 2D-[^{1}H/^{15}N]HSQC provides the correlation between $^{1}H^{N}$ and ^{15}N, one peak in the spectrum corresponds to one residue from proteins. In a peak from triple resonance spectra, correlation between ^{1}H and ^{15}N also matched to one residue from proteins (Kay *et al*., 2011). Therefore, based on peaks of 2D-[^{1}H/^{15}N]HSQC spectrum, peaks on the each spectrum can be clustered as shown in Fig. 2.

In the HN(CO)CACB scheme, only the inter-residue correlations of the form {$^{13}C^{\alpha}$(i-1), $^{13}C^{\beta}$(i-1), ^{15}N(i), $^{1}H^{N}$(i)} are observed, while both intra- and

inter-residue correlations of the form {$^{13}C^{\alpha}$(i-1), $^{13}C^{\beta}$(i-1), $^{13}C^{\alpha}$(i), $^{13}C^{\beta}$(i), ^{15}N(i), $^{1}H^{N}$(i)} are observed in the HNCACB scheme (Kay *et al*., 2011). In case of the HNCO, only $^{13}C'$(i-1), ^{15}N(i), $^{1}H^{N}$(i) correlations are observed. 3D-^{15}N-TOCSY-HSQC shows correlation between amide proton and all side-chain protons from one residue of proteins (Marion *et al*., 1989). From the 3D-^{15}N-TOCSY-HSQC spectrum, $^{1}H^{N}$-^{15}N-$^{1}H^{\alpha}$ correlations are observed. As a result of peak clustering, one peak cluster has a information of chemical shifts set {$^{1}H^{N}$(i), ^{15}N(i), $^{13}C^{\alpha}$(i), $^{13}C^{\beta}$(i), $^{13}C^{\alpha}$(i-1), $^{13}C^{\beta}$(i-1), ^{13}CO(i-1), $^{1}H^{\alpha}$(i)}.

2. Sequential assignments by linking peak clusters

Sequential backbone assignments of BldD-CTD could be accomplished by sequential linking of the peak clusters identified. As shown in Fig. 3, strip plots of the 3D-HNCACB spectra presented for residues from T109 to Q118 is sequentially linked from the correlation $^{1}H^{N}$(i)-^{15}N(i)-$^{13}C^{\alpha}$(i)-$^{13}C^{\beta}$(i)-$^{13}C^{\alpha}$(i-1)-$^{13}C^{\beta}$(i-1). As a result from sequential backbone linking procedure, the assignment of the 2D-[^{1}H/^{15}N]HSQC is shown in Fig. 4 and assigned chemical shifts of $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, ^{13}CO, $^{1}H^{\alpha}$ are summarized in Table 1. In the 2D-[^{1}H/^{15}N]HSQC spectrum (Fig. 4), the seven pairs of NH_2 signals from six glutamines and one asparagine could be clearly distinguished from the

backbone amide resonances (indicated by lines in Fig. 4). However, a few signals without labeling in Fig. 4, which might originate from the non-assigned backbone amides or from the arginine guanidine groups, could not be assigned due to the absence of corresponding signals in the 3D NMR spectra. Finally, extents of the present sequence-specific assignments of native sequence correspond to 97.6% for $^{1}H^{N}$, 90.9% for ^{15}N, 96.6% for $^{13}C^{\alpha}$, 95.3% for $^{13}C^{\beta}$, and 90.8% for $^{13}C'$ atoms. Then, total 85 of $^{1}H^{\alpha}$ atoms were unambiguously assigned based on the $^{1}H^{N}$-^{15}N-$^{1}H^{\alpha}$ correlations in 3D-^{15}N-TOCSY-HSQC spectra.

3. CSI and TALOS+ prediction of BldD-CTD

The CSI (Wishart et al., 1992; Wishart et al., 1994) method and TALOS (Cornilescu *et al*., 1999) were developed to predict a secondary structure based on chemical shifts of protein backbone atoms. Because of dependency on backbone atoms, CSI method and TALOS can also provide a NOE-independent structure prediction. Since the CSI method is a statistical technique, and the TALOS-derived ϕ/φ-values are empirical in nature, these approaches to structural information of proteins complement each other.

In the CSI method, the mark ‘1’ represents the α-helix tendency of the

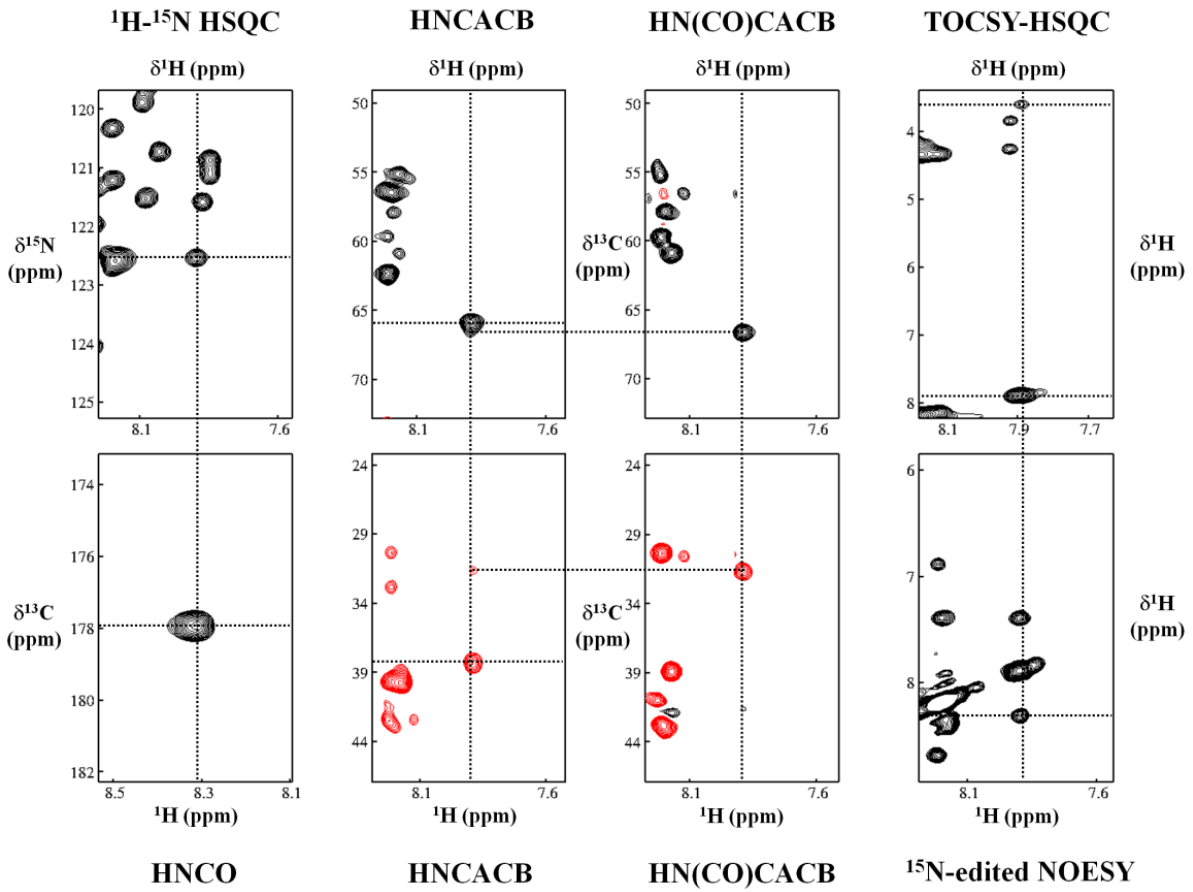

^{1}H-^{15}N HSQC
HNCACB
HN(CO)CACB
TOCSY-HSQC
δ^{1}H (ppm)
δ^{15}N (ppm)
δ^{13}C (ppm)
^{1}H (ppm)
HNCO
HNCACB
HN(CO)CACB
^{15}N-edited NOESY

Fig. 2. Verification of a peak cluster of Ile135 for backbone assignment. The 8 peaks are combined into a cluster. The X-axis of the individual spectrum represents the ^{1}H chemical shifts. The Y-axis of the 2D-[$^{1}H/^{15}N$]HSQC spectrum represent $^{15}N^{H}$ chemical shifts, and those of the 3D-^{15}N-TOCSY-HSQC and 3D-[^{15}N]-edited NOESY spectra represents ^{1}H chemical shifts. The Y-axis of the other spectra represents the ^{13}C chemical shifts of the same $^{15}N^{H}$-edited slide. Each peak at the center of the crosshair in each spectrum appears with the same ^{1}H and $^{15}N^{H}$ chemical shifts.

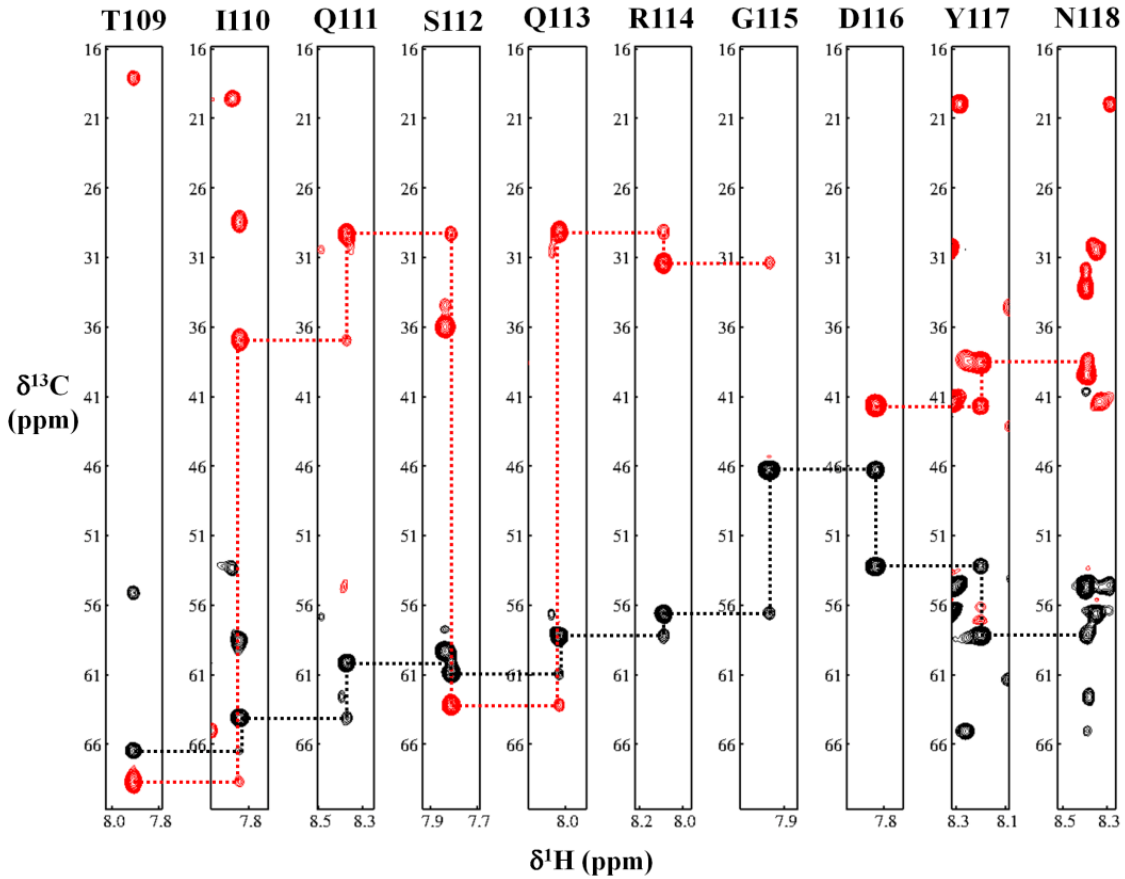

T109
I110
Q111
S112
Q113
R114
G115
D116
Y117
N118
δ13C (ppm)
δ1H (ppm)

Fig. 3. Sequential linking of peak clusters from T109 to Q118. The 10 strip plots of 3D-HNCACB are visualized for sequential linking. As a correlation of $^{1}H^{N}$(i)-^{15}N(i)-$^{13}C^{\alpha}$(i)-$^{13}C^{\beta}$(i)-$^{13}C^{\alpha}$(i-1)-$^{13}C^{\beta}$(i-1), $13C^{\alpha}$ (black dotted line) and $13C^{\beta}$ (red dotted line) are sequentially linked through the plots. The broken linking trace line shows the glycine residue (G115) which has not C^{β}.

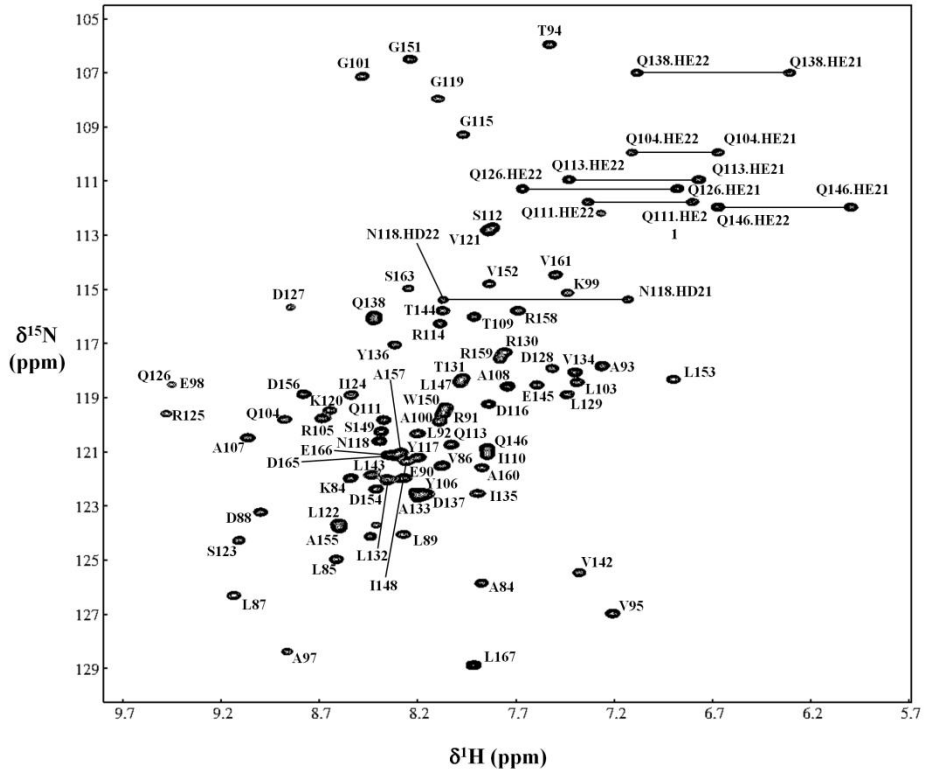


Fig. 4. 2D-[^{1}H/^{15}N]HSQC spectrum of BldD-CTD. Each correlation in the spectrum is labeled with the corresponding residue finally assigned. Side-chain signals from glutamine and asparagine residues are also labeled from assignment of 3D-[^{15}N]-edited NOESY spectrum and indicated by horizontal lines.

Table 1. Backbone assignments of BldD-CTD, measured at 295 K and pH 6.8 (NA, not available; ND, not detected).

Residue	$^{1}H^{N}$	$^{15}N^{H}$	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C'$	$^{1}H^{\alpha}$
E80	ND	ND	ND	ND	ND	ND
P81	NA	NA	ND	ND	ND	ND
P82	NA	NA	ND	ND	ND	ND
P83	NA	NA	62.78	32.35	176.3	4.401
K84	8.542	122.0	56.17	33.27	175.7	4.374
L85	8.615	125.0	54.05	43.17	175.0	4.523
V86	8.078	121.5	61.21	34.60	175.0	4.547
L87	9.136	126.3	56.55	44.23	175.8	4.940
D88	9.000	123.2	53.59	40.82	176.7	4.848
L89	8.272	124.0	57.01	40.85	180.2	4.066
E90	8.270	122.0	59.63	29.56	179.7	4.174
R91	8.075	119.6	57.31	29.56	179.9	4.172
L92	8.200	120.3	57.57	41.80	177.2	3.947
A93	7.262	117.8	54.32	18.71	178.5	4.125
T94	7.529	106.0	61.64	69.96	174.6	4.378
V95	7.206	127.0	60.54	32.39	ND	3.985
P96	NA	NA	63.85	32.71	178.1	4.268
A97	8.865	128.4	55.32	18.91	180.0	3.944
E98	9.453	118.5	59.64	28.84	177.6	4.114
K99	7.437	115.1	56.99	33.95	177.0	4.652
A100	8.096	119.9	53.18	20.23	177.5	4.144
G101	8.482	107.1	49.01	NA	ND	4.138
P102	NA	NA	65.95	31.37	179.3	4.174
L103	7.390	118.4	57.94	42.35	177.9	4.356
Q104	8.879	119.8	59.77	27.55	179.9	3.894
R105	8.688	119.8	59.64	30.42	179.0	4.141
Y106	8.214	122.6	62.28	39.71	177.8	4.171
A107	9.065	120.5	55.19	17.93	178.7	3.762
A108	7.741	118.6	55.08	18.09	180.9	4.287
T109	7.912	116.0	66.45	68.87	176.7	3.995

Table 1. (Continued)

Residue	$^{1}H^{N}$	$^{15}N^{H}$	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C'$	$^{1}H^{\alpha}$
I110	7.844	121.1	64.15	36.94	178.6	3.815
Q111	8.373	119.8	59.86	29.37	178.3	3.839
S112	7.813	112.7	60.92	63.24	176.8	4.333
Q113	8.029	120.7	58.11	29.19	177.9	4.194
R114	8.088	116.3	56.42	31.42	177.0	4.398
G115	7.968	109.3	46.27	NA	173.7	3.880
D116	7.838	119.2	53.17	41.64	175.4	4.715
Y117	8.199	121.2	58.13	38.59	176.9	4.527
N118	8.391	120.6	54.38	39.51	175.9	4.603
G119	8.097	108.0	45.93	NA	173.9	3.929
K120	8.647	119.5	57.71	34.35	176.2	4.489
V121	7.838	112.8	59.26	35.95	174.9	5.330
L122	8.604	123.7	53.88	47.50	175.3	4.789
S123	9.110	124.3	58.85	63.51	174.3	5.052
I124	8.540	118.9	59.60	41.60	174.2	4.780
R125	9.481	119.6	54.51	33.32	178.5	4.865
Q126	9.345	122.8	60.53	28.28	179.0	3.818
D127	8.848	115.7	57.36	40.48	178.8	4.360
D128	7.515	117.9	58.08	41.64	177.9	4.612
L129	7.441	118.9	58.13	41.40	178.3	3.912
R130	7.758	117.3	60.29	29.82	178.9	3.910
T131	7.975	118.3	67.00	68.76	176.6	3.924
L132	8.354	122.0	57.82	43.09	177.8	3.809
A133	8.197	122.6	56.51	17.16	179.2	3.892
V134	7.401	118.1	66.59	31.69	180.4	3.861
I135	7.894	122.5	65.84	38.39	177.9	3.604
Y136	8.317	117.1	60.76	39.03	174.7	4.131
D137	8.173	122.6	55.11	39.81	174.9	4.350
Q138	8.422	116.0	53.51	34.38	174.9	4.719
S139	8.422	116.1	56.44	62.81	ND	4.783

Table 1. (Continued)

Residue	$^{1}H^{N}$	$^{15}N^{H}$	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C'$	$^{1}H^{\alpha}$
P140	NA	NA	66.50	31.78	179.2	4.036
S141	ND	ND	62.37	62.29	176.1	4.838
V142	7.378	125.5	66.43	32.02	178.8	3.561
L143	8.438	121.9	57.02	41.26	178.8	3.899
T144	8.074	115.8	69.06	67.99	175.6	3.321
E145	7.593	118.5	59.23	28.92	179.3	3.841
Q146	7.845	120.9	58.39	28.37	177.4	3.100
L147	7.982	118.4	58.41	40.97	180.2	3.473
I148	8.258	121.4	65.05	38.35	180.9	4.160
S149	8.386	120.3	62.54	62.84	177.0	4.273
W150	8.061	119.4	55.76	31.26	176.8	4.926
G151	8.239	106.5	45.40	NA	174.7	4.167
V152	7.835	114.8	63.11	32.85	174.0	4.197
L153	6.897	118.3	52.63	47.16	175.0	4.911
D154	8.412	122.4	54.38	43.23	176.4	4.603
A155	8.600	123.8	54.73	18.75	179.1	4.054
D156	8.784	118.9	56.28	40.91	174.9	4.513
A157	8.287	121.0	54.39	19.92	178.8	4.001
R158	7.688	115.8	59.23	30.11	178.3	3.924
R159	7.781	117.6	57.69	30.24	177.2	4.206
A160	7.872	121.6	53.26	19.56	177.8	4.279
V161	7.500	114.5	62.04	32.78	175.6	4.141
A162	7.874	125.8	52.77	19.51	177.7	4.321
S163	8.248	115.0	58.39	63.92	ND	4.396
H164	7.971	125.4	56.32	30.31	174.9	4.666
D165	8.322	121.1	54.58	41.26	175.9	4.603
E166	8.344	121.1	56.37	30.37	175.5	4.317
L167	7.916	128.9	57.15	43.45	ND	4.165

atom of the residue, while the mark '-1' represents β-strand tendency. The mark '0' shows that residue does not have a strand or a helical tendency, of which the chemical shift is within the reference value range.

The TALOS is a new database system for the prediction of ϕ and φ backbone torsion angles using a combination of chemical shift assignments for a given protein sequence (Cornilescu *et al*., 1999). The TALOS output for ϕ and φ angles of the center residue in each string consists of the average of corresponding angles in the 10 strings in the database with the highest degree of similarity. The TALOS+ is enhanced version of TALOS which further enhances the prediction rate to 88.5%, without increasing the error rate from addition of a two-layer neural network filter to the database fragment selection process (Shen et al., 2009).

All of the assigned chemical shift data sets were applied to CSI and TALOS+ analysis to predict the secondary structure elements of BldD-CTD (Fig. 5). The determined secondary structure by combining the CSI and TALOS+ results indicated that BldD-CTD is predominantly α-helical, similar to the previous far-UV CD results (Lee et al., 2007). However, the present results additionally revealed the presence of two short β-strands, therby forming a βααβααα topology. The two β-strands are expected to form a β-sheet since

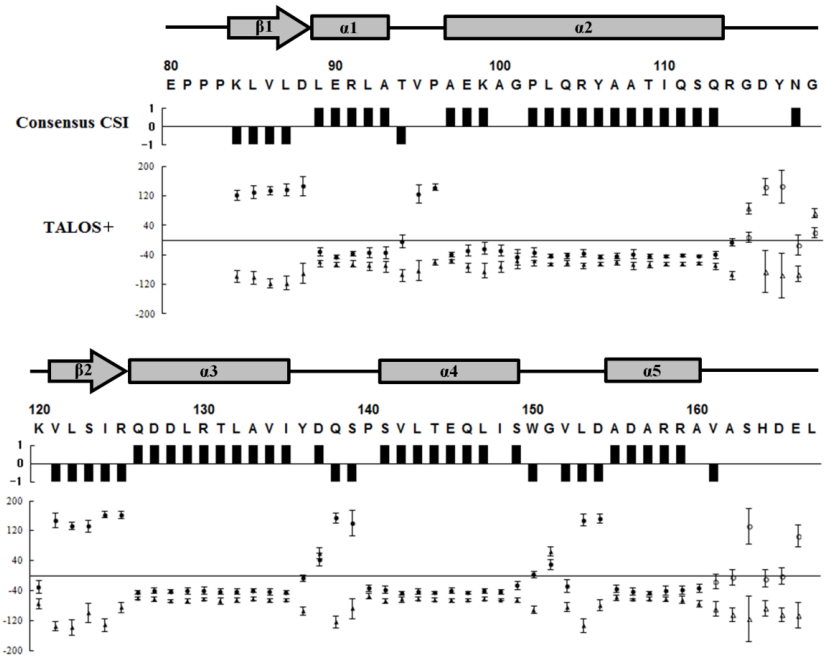


Fig. 5. Secondary structure prediction from CSI methods and TALOS+. In the CSI results, the mark "1" represents the α-helical tendency of the residue (downfield shifts of $^{13}C^{\alpha}$ or $^{13}C'$ resonances, and upfield shifts of $^{13}C^{\beta}$ or $^{1}H^{\alpha}$ resonances from their reference value ranges), while "-1" represents the opposite pattern (β-strand tendency). The chemical shift within the reference value range was marked as a "0". The length of error bars with the TALOS+-predicted backbone dihedral angles, ϕ (triangles) and φ (rectangles), indicate the standard deviation from the average. Predicted secondary structure elements are indicated by boxes for α-helices and arrows for β-strands, along the amino acid sequence.

their lengths are almost the same.

4. Verification of peak clusters for side-chain assignments

Based on backbone assignments, side-chain assignment was performed through analyses of 3D-HCCH-COSY (Ikura *et al*., 1991) and 3D-HCCH-TOCSY (Olejniczak *et al*., 1992). In 3D-HCCH-COSY and 3D-HCCH-TOCSY, magnetisation is transferred from the side-chain hydrogen nuclei to their attached ^{13}C nuclei. While only the hydrogen resonances of the own and within three covalent bond in 3D-HCCH-COSY, all side-chain hydrogen resonances of residue are visibible in 3D-HCCH-TOCSY. 2D-[$^{1}H/^{13}C$]HSQC spectrum which is correlation between ^{1}H and ^{13}C was also used for verification of side-chain assignment. From the backbone assignment of BldD-CTD, chemical shifts of $^{13}C^{\alpha}(i)$, $^{1}H^{\alpha}(i)$, $^{13}C^{\beta}(i)$ were used for peak clustering of 3D-HCCH-COSY, HCCH-TOCY and 2D-[$^{1}H/^{13}C$]HSQC as shown in Fig. 6. Using the difference of correlation between HCCH-TOCSY and HCCH-COSY, 98.3 % of side-chain including chemical shifts of $^{13}C^{\alpha}$, $^{1}H^{\alpha}$, $^{13}C^{\beta}$ were assigned (Fig. 7).

5. NOESY cross-peak assignment of BldD-CTD

The NOESY allows correlating nuclei through space (distance smaller than 5Å). By measuring cross-peak intensity or volume, distance information can be

acquired (Wüthrich K, 1986). To avoid overlap problem of NOE cross-peak measured from BldD-CTD, 3D-[^{15}N]-edited NOESY and 3D-[^{13}C]-edited NOESY spectra were used for NOE assignment (Marion *et al*., 1989). Based on assigned chemical shifts, NOE cross-peaks in each diagonal peak were assigned (Fig. 8). As information of distances is already known for intra-residues, intensities of intra-residual cross-peaks are used as distance standards (Wüthrich K, 1986). To calculate 3D protein structure, the set of lower and upper distance limits among protons is required. As the sum of van der Waals radii for two protons (about 1.8 Å) used as a lower distance limit in general, upper distance limits are critical for determination of protein structure. From these, each upper limit distance for strong, medium, and weak cross-peak was determined to 2.5-3.0 Å, 4.0 Å, and 5.0 Å. Among them, long-range NOEs, which are correlations between atoms separated by five or more residues along the protein sequence, provide important 3D-structural information for protein (Wüthrich K, 1986). The upper distance limits for long-range NOEs of BldD-CTD listed in Table 2.

6. 3D-Structure determination using CYANA

Using the upper distance limits from assignment of NOE cross-peaks, initial structure of BldD-CTD calculated by dihedral torsion angle dynamics

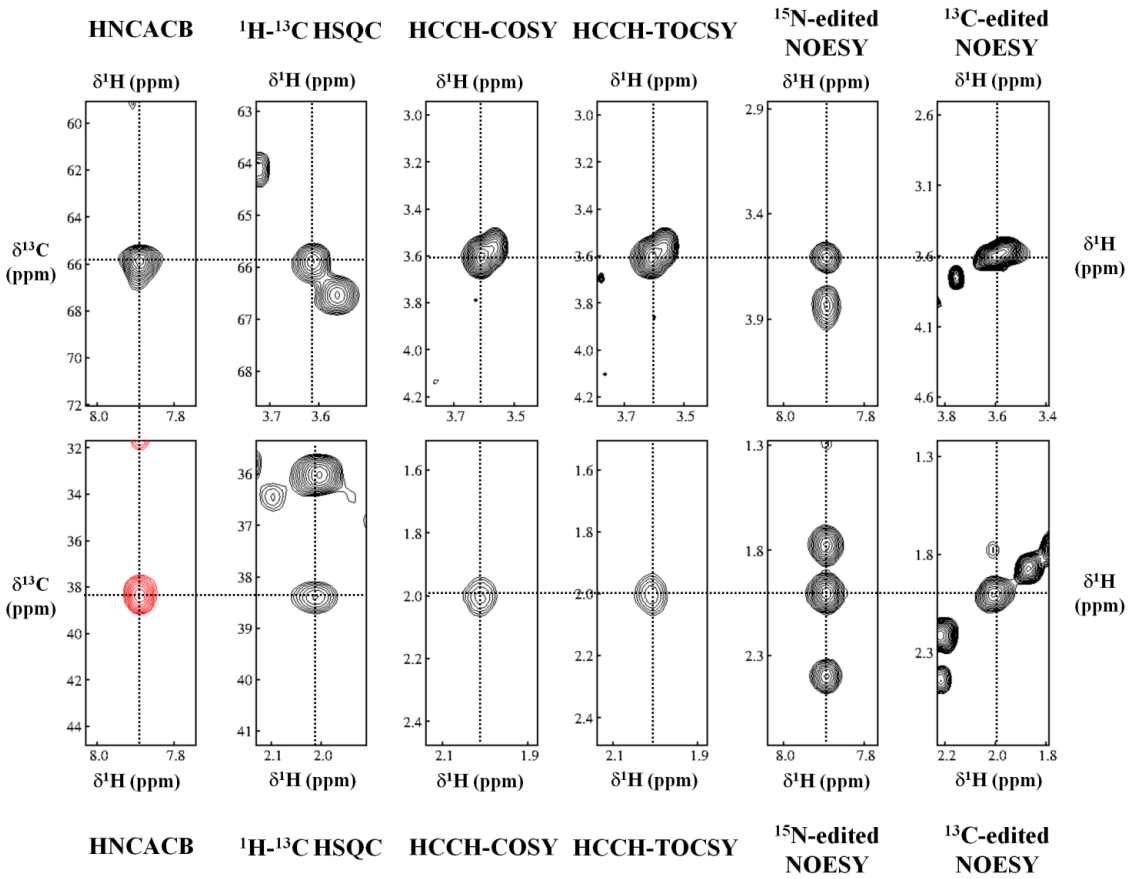

HNCACB
¹H-¹³C HSQC
HCCH-COSY
HCCH-TOCSY
¹⁵N-edited NOESY
¹³C-edited NOESY
δ¹H (ppm)
δ¹³C (ppm)
δ¹H (ppm)
δ¹³C (ppm)
δ¹H (ppm)
HNCACB
¹H-¹³C HSQC
HCCH-COSY
HCCH-TOCSY
¹⁵N-edited NOESY
¹³C-edited NOESY

Fig. 6. Verification of a peak cluster of Ile135 for side-chain assignment. The 6 peaks for $^{13}C^{\alpha}$ (upper line) and 6 peaks for $^{13}C^{\beta}$ (bottom line) are combined into a cluster. The X-axis of the individual spectrum represents the ^{1}H chemical shifts. The Y-axis of the HNCACB and 2D-[$^{1}H/^{13}C$]HSQC spectrum represent ^{13}C chemical shifts, and those of the other spectra represents $^{1}H^{\alpha}$ (upper line) or $^{1}H^{\beta}$ (bottom line) chemical shifts. 3D-HNCACB and 3D-[^{15}N]-edited NOESY have the same $^{15}N^{H}$-edited slide, and other spectra except 2D-[$^{1}H/^{13}C$]HSQC have the same $^{13}C^{\alpha}$- (upper line) or $^{13}C^{\beta}$- (bottom line) edited slide.

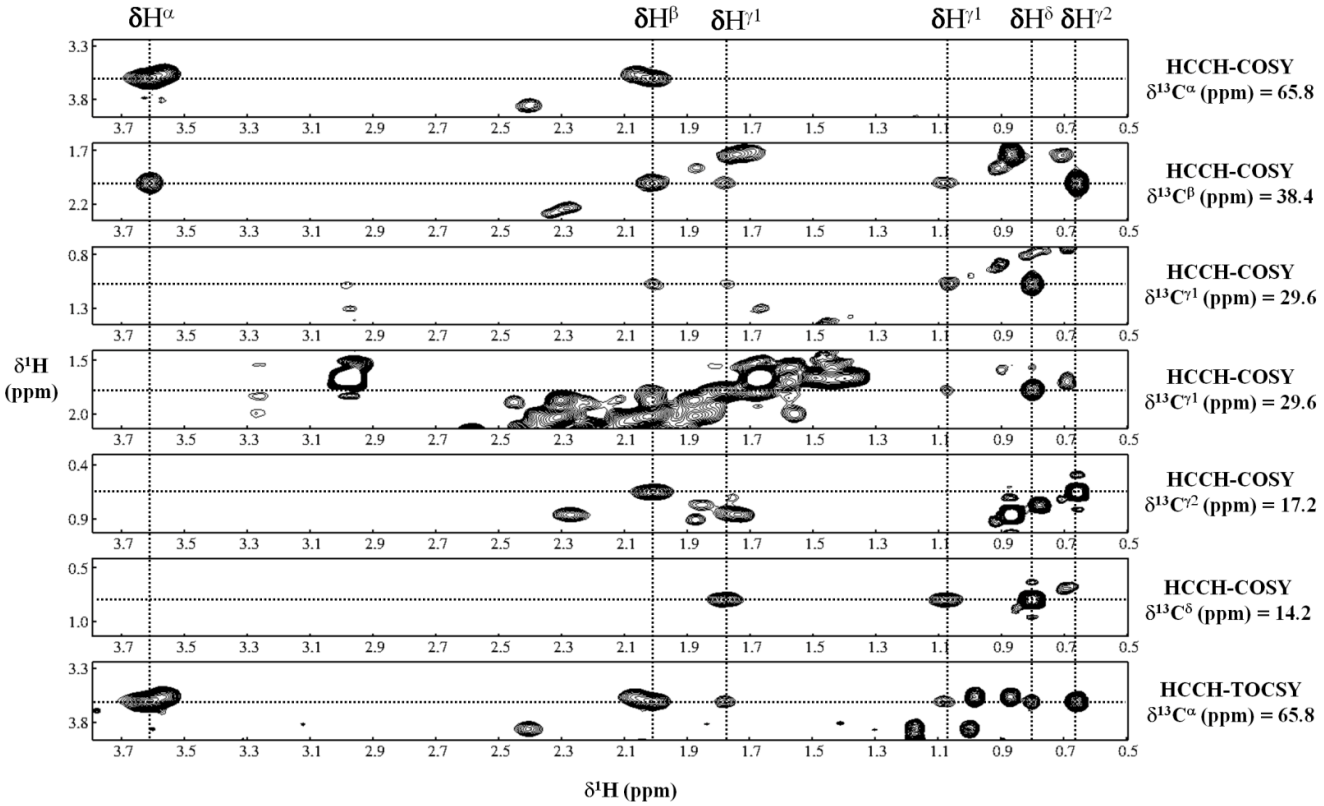

δH^α
δH^β
δH^γ1
δH^γ1
δH^δ
δH^γ2
HCCH-COSY
δ13Cα (ppm) = 65.8
HCCH-COSY
δ13Cβ (ppm) = 38.4
HCCH-COSY
δ13Cγ1 (ppm) = 29.6
HCCH-COSY
δ13Cγ1 (ppm) = 29.6
HCCH-COSY
δ13Cγ2 (ppm) = 17.2
HCCH-COSY
δ13Cδ (ppm) = 14.2
HCCH-TOCSY
δ13Cα (ppm) = 65.8
δ1H (ppm)
δ1H (ppm)

Fig. 7. Side-chain assignment of Ile35 using HCCH-COSY and HCCH-TOCSY. Each spectrum is labeled on its right side. ^{13}C chemical shifts for editing spectrum are also labeled on its right side. Above vertical dotted lines ^{1}H chemical shifts correlated among spectra are labeled. Horizontal dotted lines represent the ^{1}H chemical shifts correlated in each spectrum.

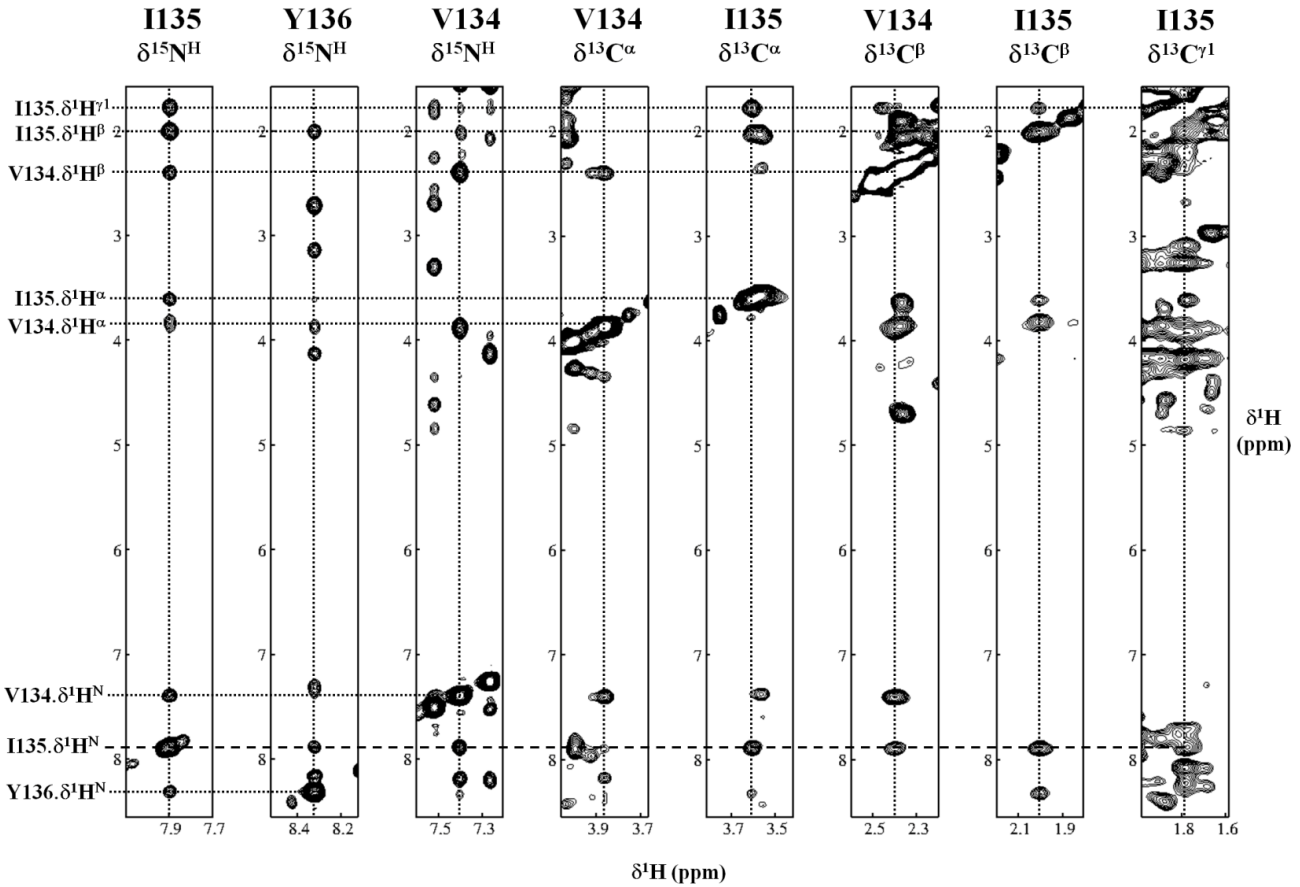

I135 δ15NH
Y136 δ15NH
V134 δ15NH
V134 δ13Cα
I135 δ13Cα
V134 δ13Cβ
I135 δ13Cβ
I135 δ13Cγ1
I135.δ1Hγ1
I135.δ1Hβ
V134.δ1Hβ
I135.δ1Hα
V134.δ1Hα
V134.δ1HN
I135.δ1HN
Y136.δ1HN
δ1H (ppm)
δ1H (ppm)

Fig. 8. NOE cross-peaks assignment of Ile135 on 3D-[^{15}N]-edited NOESY. Above vertical dotted lines in each spectrum, corresponding residue and ^{1}H chemical shift are labeled. Among the 8 strip plots, 3 plots on the left side are 3D-[^{15}N]-edited NOESY and other plots are 3D-[^{13}C]-edited NOESY. From the correlation among spectra, assignments of NOE cross-peaks for Ile135 are labeled on the left side of spectrum corresponding to Ile135.

Table 2. Upper distance limits for long-range NOEs. The nomenclature of atoms in amino acid residues follows the IUPAC recommendations. 'A', 'B', 'G', 'D', and 'E' represent 'α', 'β', 'γ', 'δ', and 'ε', respectively. 'Q' represents pseudoatom (Markley et al, 1998). 'Upl' represents upper distance limit.

Residue	Atom	Residue	Atom	Upl (Å)	Residue	Atom	Residue	Atom	Upl (Å)
83 PRO	HB2	157 ALA	HA	5.5	95 VAL	QG2	100 ALA	QB	5.5
83 PRO	QG	124 ILE	H	6	95 VAL	QG1	100 ALA	QB	6
84 LYS	QG	120 LYS	HA	6.5	95 VAL	QG2	150 TRP	HE1	6.5
86 VAL	QG2	121 VAL	HA	6	99 LYS	HD3	150 TRP	HE1	5.5
86 VAL	HA	123 SER	HA	5	99 LYS	HD2	150 TRP	HE1	5.5
86 VAL	QG1	123 SER	HA	6	103 LEU	QD2	108 ALA	H	6.5
86 VAL	QG1	123 SER	QB	6	103 LEU	HA	132 LEU	QD1	6
86 VAL	QG2	123 SER	HA	5	103 LEU	HB3	136 TYR	HE1	5.5
86 VAL	HA	124 ILE	H	5	103 LEU	HB3	136 TYR	HE2	5.5
87 LEU	H	122 LEU	H	5	103 LEU	HB2	136 TYR	HE1	5.5
88 ASP	H	152 VAL	HA	5	103 LEU	HB2	136 TYR	HE2	5.5
88 ASP	H	152 VAL	QG2	5.5	103 LEU	QD1	136 TYR	HE1	6.5
88 ASP	HB2	152 VAL	HA	5	103 LEU	QD1	136 TYR	HE2	6.5
88 ASP	HB3	152 VAL	HA	4	103 LEU	QD2	136 TYR	HE1	6.5
89 LEU	QD2	107 ALA	QB	5	103 LEU	QD2	136 TYR	HE2	6.5
89 LEU	QD1	111 GLN	HE22	6.5	103 LEU	QD1	147 LEU	QD1	6
89 LEU	QD1	111 GLN	HE21	6	103 LEU	QD1	147 LEU	QD2	6
89 LEU	H	121 VAL	QG1	6	103 LEU	QD1	150 TRP	HZ3	6.5
89 LEU	H	121 VAL	HA	4	103 LEU	QD2	150 TRP	HZ3	6.5
89 LEU	QD1	121 VAL	HA	6.5	106 TYR	HB3	132 LEU	QD1	6
89 LEU	QD2	121 VAL	HA	6.5	106 TYR	HB2	132 LEU	QD1	6
89 LEU	QD1	122 LEU	H	6	106 TYR	QE	128 ASP	HB2	6.5
89 LEU	QD2	122 LEU	H	6.5	107 ALA	H	132 LEU	QD1	6.5
89 LEU	HG	121 VAL	HA	5	107 ALA	HA	132 LEU	QD1	6
92 LEU	QD1	104 GLN	H	6	111 GLN	HA	116 ASP	HB2	5.5
92 LEU	QD1	104 GLN	HE21	6	111 GLN	HA	116 ASP	HB3	5.5
92 LEU	QD1	104 GLN	HE22	6	111 GLN	HE21	117 TYR	HA	6
93 ALA	HA	104 GLN	HE22	5	111 GLN	HE22	117 TYR	HA	6
93 ALA	HA	104 GLN	HE21	6	111 GLN	HE22	122 LEU	QD1	6.5
95 VAL	H	100 ALA	QB	6	111 GLN	HG2	122 LEU	QD1	6
95 VAL	HB	100 ALA	QB	5	111 GLN	HG3	122 LEU	QD1	6
95 VAL	QG1	100 ALA	H	6	133 ALA	H	143 LEU	QD2	6
95 VAL	QG2	100 ALA	H	6	137 ASP	H	143 LEU	QD1	5.5
95 VAL	QG2	100 ALA	HA	6.5	148 ILE	QD1	158 ARG	H	6

algorithm DYANA embedded in CYANA 2.1 program (Herrmann *et al*., 2002). The initial ensemble structure of BldD-CTD showed average backbone RMSD of 1.73 ± 0.48 Å, which fulfilled the quality required for reliable structure (Güntert, 2004) (Fig. 9A). As shown in Fig. 9B, representative initial structure of BldD-CTD had four α-helices and a weak helical region (denoted as η1).

To obtain more NOE assignments which were not identified by manual assignment, and higher resolution of 3D-structure, CANDID module for automated NOE assignment was used for calculation of structure. As the CANDID module uses algorithms of ambiguous peak assignment and network anchoring during seven cycles which uses NOE assignment and calculated structure of the previous cycle as templates, 3D-structure of high quality can be calculated (Güntert, 2004). As expected that, fine 3D-structure of BldD-CTD (Average backbone RMSD = 0.60 ± 0.11 Å) was calculated from combination of CANDID and DYANA (Fig. 10A). It also had four α-helices and a weak helical region η1, which was identical to initial structure from manual NOE assignment (Fig. 10B). In addition, although RMSD between two representative structures over 73 atoms of C^{α} (P83-R114 and K120-A160) was 2.330 Å, as ensemble structure from manual assignment had higher RMSD than that from automated assignment and secondary structure regions were identical, two

structures from different methods are almost same (Fig. 11). For reliable structure calculation from automated NOE assignment, there are six criteria as follows (Güntert, 2004):

1. Average CYANA target function value of cycle 1 below 250 $Å^2$.
2. Average final CYANA target function value below 10 $Å^2$.
3. Less than 20% unassigned NOEs.
4. Less than 20% discarded long-range NOEs.
5. RMSD value in cycle 1 below 3 Å.
6. RMSD between the mean structures of the first and last cycle below 3 Å.

As shown in Table 3, all criteria except criterion 4 were fulfilled for automated structure calculation of BldD-CTD. The percentage of discarded long-range NOEs cannot be calculated readily outside the CYANA program because it requires knowledge of the possible assignments. In this case, an overall percentage of unused cross-peaks less than 15% can be used as an alternative criterion (Güntert, 2004). In the cycle 7, overall proportion of unassigned cross-peaks was only 8.6 %, which satisfied alternative criterion 4. The β-sheet predicted from CSI method and TALOS (K84-L87 and V121-I124) was not clearly seen in both calculated structures (Fig. 9 and 10). But, because long-range NOEs are detected in these regions (Table 2) and they are not

refined structures, it is so reasonable that a β-sheet exists in these regions.

7. Refinement of the structure using CNS

For improving efficiency of three dimensional protein structure calcuation, simplification of the nonbonded interactions was applied to the process of calculation. This causes unreal treatment of electrostatic and van der Waals (vdW) interactions, followed by unsatisfactory quality indices from validation programs. Therefore, refinement process is essential for bringing the protein structures closer to physical reality (Linge *et al*., 2003).

To refine the structure of the BldD-CTD, CNS program (Brünger *et al*., 1998) using RECOORD scripts (Nederveen *et al*., 2005) were used for recalculation of the BldD-CTD structure. As a result, final 20 structures of the BldD-CTD were calculated (Fig. 12) and validated using PSVS 1.5 (Bhattacharya *et al*., 2007) software suites (as summarized in Table 4).

8. Overall structure of BldD-CTD

The solution NMR structure of BldD-CTD (Fig. 12 and 13) consists of four α-helices, a two-stranded anti-parallel β-sheet and a C-terminal weak helical region (designated as η1) with topology of βααβααη. Helices α1 (L89-

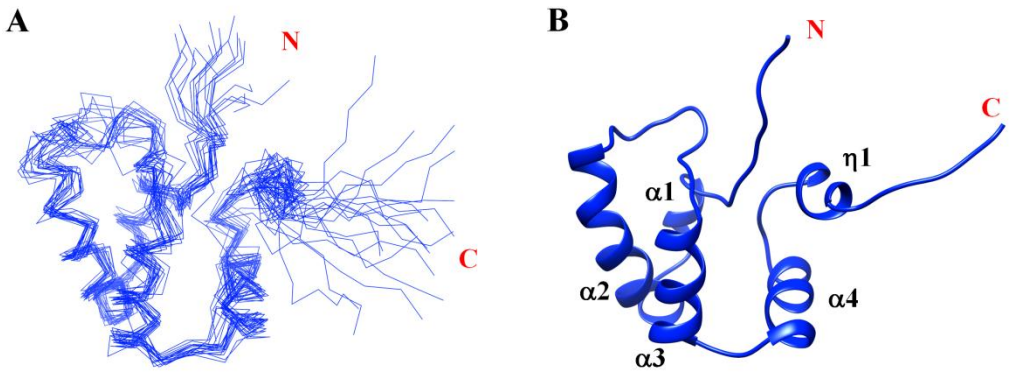


Fig. 9. Initial BldD-CTD structure calculated from manual assignment of NOE cross-peaks. (A) 20 structures of BldD-CTD calculated from manual assignment of NOE cross-peaks are superimposed over equivalent C^{α} atoms in the ordered regions (P83-R114 and K120-A160). (B) Cartoon representation for representative structure of (A). Secondary structure elements are labeled. 'N' and 'C' represent N-terminus and C-terminus, respectively.

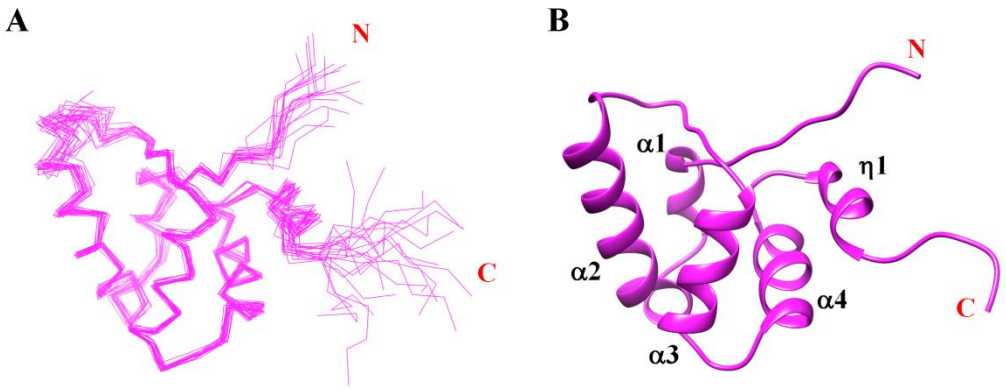


Fig. 10. BldD-CTD structure calculated from automated assignment of NOE cross-peaks. (A) 20 structures of BldD-CTD calculated from automated assignment of NOE cross-peaks are superimposed over equivalent C^{α} atoms in the ordered regions (P83-R114 and K120-A160). (B) Cartoon representation for representative structure of (A). Secondary structure elements are labeled. 'N' and 'C' represent N-terminus and C-terminus, respectively.

Table 3. Statistics of automated BldD-CTD structure calculation

Cycle	1	2	3	4	5	6	7	
Peaks:								
selected	3060	3060	3060	3060	3060	3060	3060	
assigned	2805	2845	2833	2833	2834	2830	2826	
unassigned	255	2215	227	227	226	230	234	
with diagonal assignment	210	212	212	212	212	212	212	
Cross peaks:								
with off-diagonal assignment	2595	2633	2621	2621	2622	2618	2614	
with unique assignment	985	1652	1857	1956	2088	2151	2172	
with short-range assignment [i-j] <=1	2023	2002	1978	1969	1950	1933	1934	
with medium range assignment 1<[i-j]<5	387	408	388	389	387	394	388	
with long-range assignment [i-j] >=5	185	223	255	263	285	291	292	
Upper distance limits:								
total	1718	1604	1523	1504	1468	1447	1467	1510
short-range [i-j] <=1	1148	1025	953	932	891	866	788	820
medium-range 1<[i-j]<5	523	511	341	337	329	332	359	363
long range [i-j] >=5	47	68	229	235	248	249	320	327
Average assignments/constraint	4.21	2.23	1.46	1.42	1.32	1.28	1.00	1.00
Average target function value	2.60	1.93	4.29	0.59	0.42	0.35	0.36	0.24
RMSD (residues 5-83):								
Average backbone RMSD to mean	1.45	0.95	0.70	0.75	0.70	0.72	0.61	0.60
Average heavy atom RMSD to mean	2.00	1.52	1.25	1.25	1.23	1.25	1.16	1.13

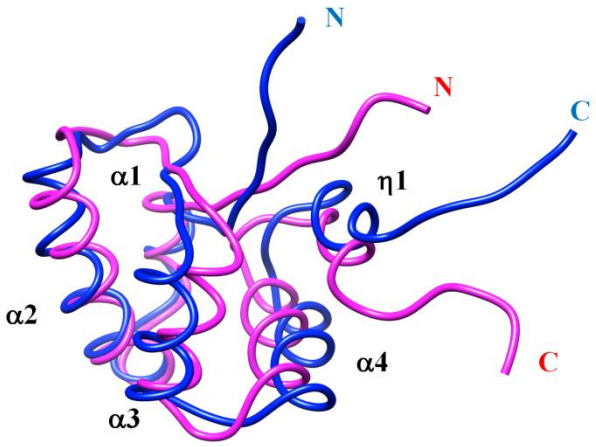


Fig. 11. Superimposed representative structures calculated from manual (blue) and automated (magenta) assignment of NOE cross-peaks. The representative structures from both ensemble structures with manual and automated assignment are superimposed over equivalent C^{α} atoms in the ordered regions (P83-R114 and K120-A160). Secondary structure elements are labeled. 'N' and 'C' represent N-terminus and C-terminus, respectively.

L92), α2 (A97-R114), α3 (L129-136) and α4 (P140-W150) form the core frame and a β-sheet composed of strand β1 (L85-D88) and β2 (V121-I124) is packed against the one side of helix bundle (Fig. 13). The overall conformation of BldD-CTD is stabilized mainly through the interaction of interior or semi-interior hydrophobic residues (L85 and L87 from β1, L89 and L92 from α1, L103 and Y106 from α2, L122 and I124 from β2, L129, L132, I135, and Y136 from α3, L143, L147, and W150 from α4, and V152 and L153 from the α4-η1 loop) maintains the main framework (Fig. 14). As seen in Fig. 12, all regions of BldD-CTD are well-ordered except for its N-terminal linker for BldD-NTD, a β-turn between helix α2 and strand β2, and a C-terminal tail. The helix η1 (A155-A161) adjacent to C-terminus shows a weakly ordered but dynamic property in its ensemble structure (Fig. 12). It appears as a hybrid of a turn of 3_{10}-helix (A155-A157) and a turn of α-helix (R158-A161) (Fig. 15).

To verify whether this helix is an erroneous structure calculated from wrong assignment, the NOE cross-peaks in this region were analyzed. As shown in Fig. 16, the region from A155 to A157 shows NOEs of $d_{\alpha N}(i,i+2)$ cross-peaks, which are typical NOEs for 3_{10}-helix,. But, NOEs of $d_{\alpha N}(i,i+4)$ for α-helix were not detected in the region from R158 to A161. Since NOEs of $d_{\alpha N}(i,i+4)$ for α-helix is very weak (Wüthrich K, 1986) and a turn of α-helix in

helix η1 is close to severely dynamic C-terminus (Fig. 12), the correlations for α-helical region of η1 might be undetectable. Although the NOEs for this region were not detected sequentially, residue R159 positioned in α-helical region of η1 showed a correlation of $d_{\alpha N}(i,i+3)$ and a strong correlations of $d_{\alpha\beta}(i,i+3)$, both of which are NOEs of α-helix (Fig. 17). Thus, it can be mentioned that this region (R158-A161) has a α-helical property.

In the [^{1}H,^{15}N]-heteronuclear NOE experiments, the determined NOE values in the η1 region were relatively low to those in other helices but higher than other disordered regions (Fig. 18), which verified ordered but dynamic property of η1. In spite of dynamic property of η1, its spatial location is maintained by loose hydrophobic packing composed of A155 and A157 from the η1, L85 from the β1, I148 from the α4, and L153 from the α4-η1 loop (Fig. 14). Thus, η1 is seemed to slightly restrict dynamic feature of C-terminal region in the BldD-CTD structure.

9. Structural comparison between BldD-NTD and BldD-CTD

BldD-NTD adopts a compact globular structure composed of four α-helices, which is packed through hydrophobic interaction among helices. Helix α2 and α3 constitute a canonical helix-turn-helix fold and helix α1 forms

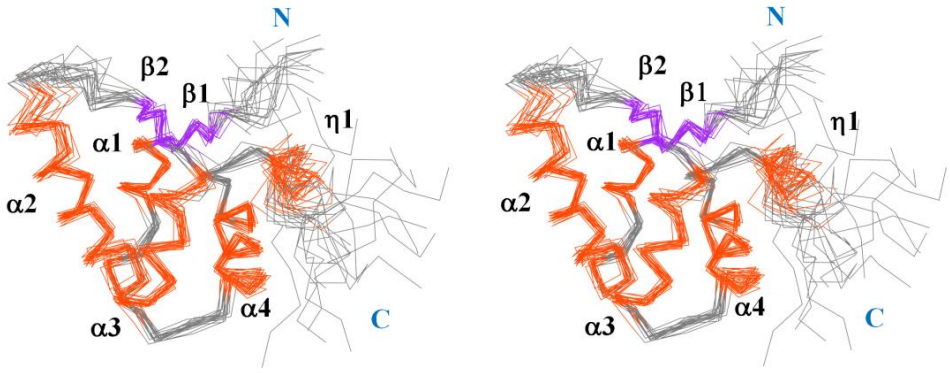


Fig. 12. Stereoview of the BldD-CTD ensemble structure from refinement process. 20 final structures of BldD-CTD are superimposed over equivalent C^{α} atoms in the ordered regions (P83-R114 and K120-A160). Secondary structure elements are labeled and colored orange for α- and η-helices and purple for β-strands.

Table 4. Summary of NMR and Structural Statistics for the final 20 conformers of the BldD-CTD[a]

Completeness of resonance assignments	
Backbone (%)	97.6
Side chain (%)	98.3
Restraint statistics	
Distance Restraints	
Total	1382
Intra-residue (i = j)	311
Sequential (|i – j| = 1)	381
Medium range (1 < |i – j| < 5)	363
Long range (|i – j| ≥ 5)	327
Distance constraints per residue	17.3
Dihedral-angle constraints	138
ϕ	69
ψ	69
Number of restraints per residue	19
Number of long-range restraints per residue	4.1
CYANA targat function (Å^2)	0.24
Residual restraint violations	
Average number of distance violations per structure	
0.1–0.2 Å	1.15
0.2–0.5 Å	0.1
>0.5 Å	0
Average RMS distance violation/restraint (Å)	0.00
Maximum distance violation (Å)	0.21
Average number of dihedral angle violation per structure	
1-10°	1.3
>10°	0
Average RMS of dihedral angle violation/constraint	0.18
Maximum dihedral angle violation (degree)	2.6

[a]Calculated using the PSVS 1.5 program (Bhattacharya *et al*., 2007).

Table 4. (Continued)

RMSD from average coordinates (Å)[b]		
Backbone atoms	0.7	
Heavy atoms	1.2	
Procheck Ramachandran statistics		
Most favored regions (%)	96.3	
Additional allowed regions (%)	3.7	
Generously allowed regions (%)	0.0	
Disallowed regions (%)	0.0	
Global quality scores (Raw/Z-score)		
Procheck (phi-psi)	0.13	0.83
Procheck (all)	0.03	0.18
Verify3D	0.41	-0.80
ProsaII	0.80	0.62
MolProbity clash	11.16	-0.39
RPF scores[c]		
Recall	0.953	
Precision	0.871	
F-measure	0.910	
DP-score	0.791	

[b]Analyzed in the ordered [S(phi) + S(psi) > 1.8] regions (P83-R114 and K120-A160).

[c]RPF scores reflect the goodness-of-fit of the final ensemble of structures.

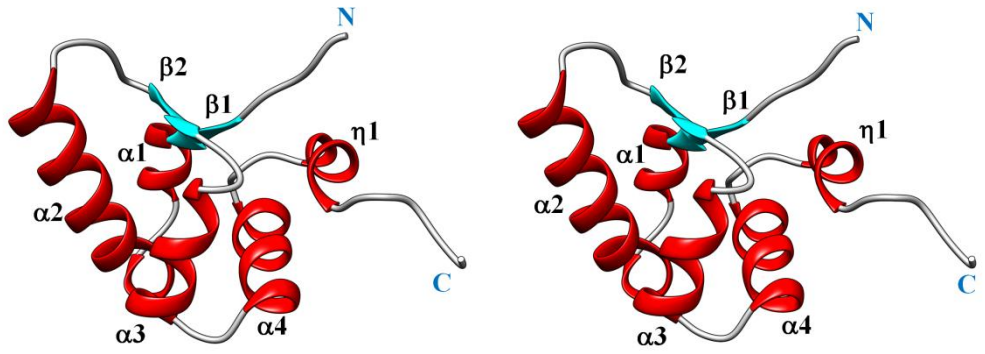


Fig. 13. Stereoview of cartoon representation for the representative BldD-CTD structure out of 20 final structures. Secondary structure elements are labeled and colored red for α- and η-helices and cyan for β-strands.

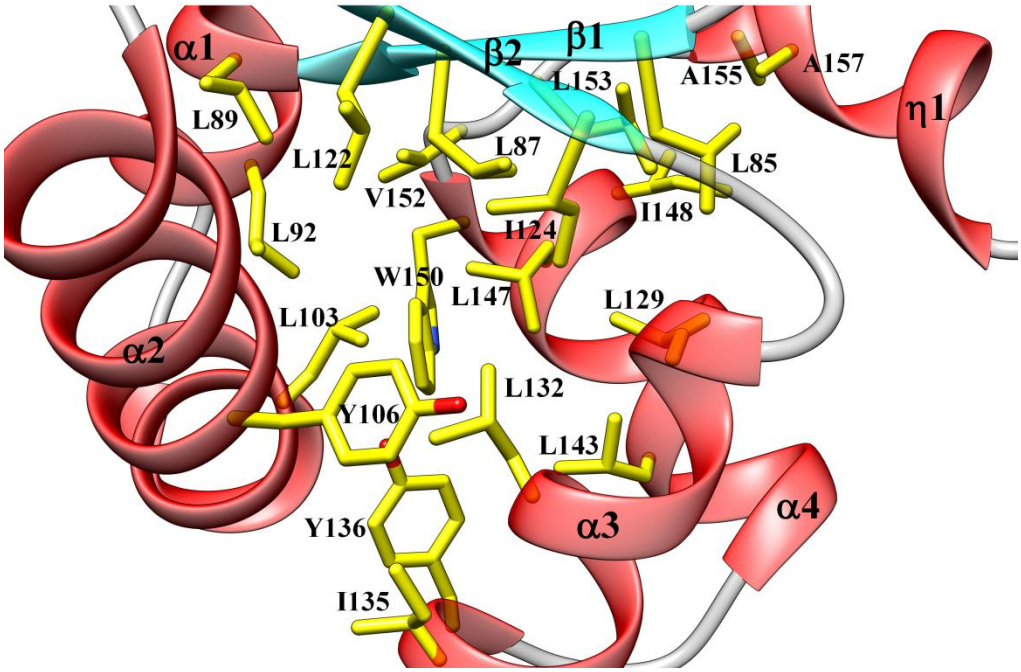


Fig. 14. Hydrophobic core formation of the BldD-CTD structure. Buried or partly buried hydrophobic side-chains are represented as yellow sticks with heteroatoms colored red for oxygen and blue for nitrogen. Ribbons of secondary structure elements are colored red for helices (α1-α4 and η1) and cyan for β-strands (β1 and β2).

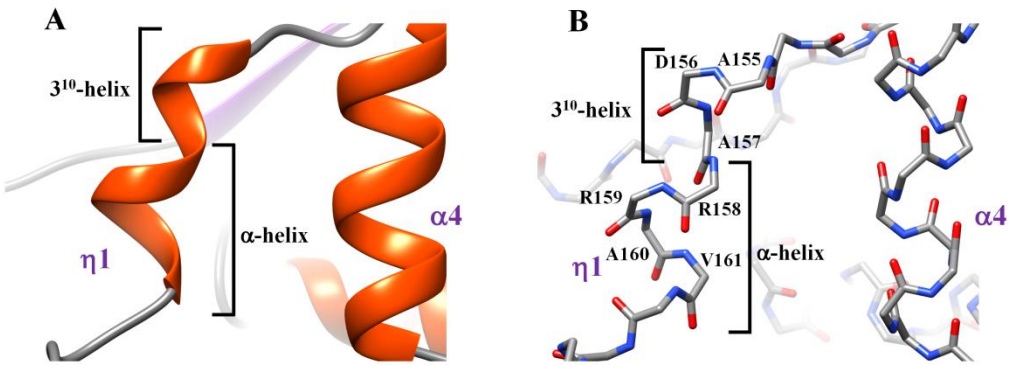


Fig. 15. Hybrid helix η1 of the BldD-CTD structure. (A) cartoon representation of helix η1. (B) Representation of backbone atoms for helix η1. Heteroatoms are colored red and blue for oxygen and nitrogen, respectively. Individual residues are also labeled. Secondary structure elements are labeled and each region for 3_{10}-helix and a-helix are indicated in (A) and (B).

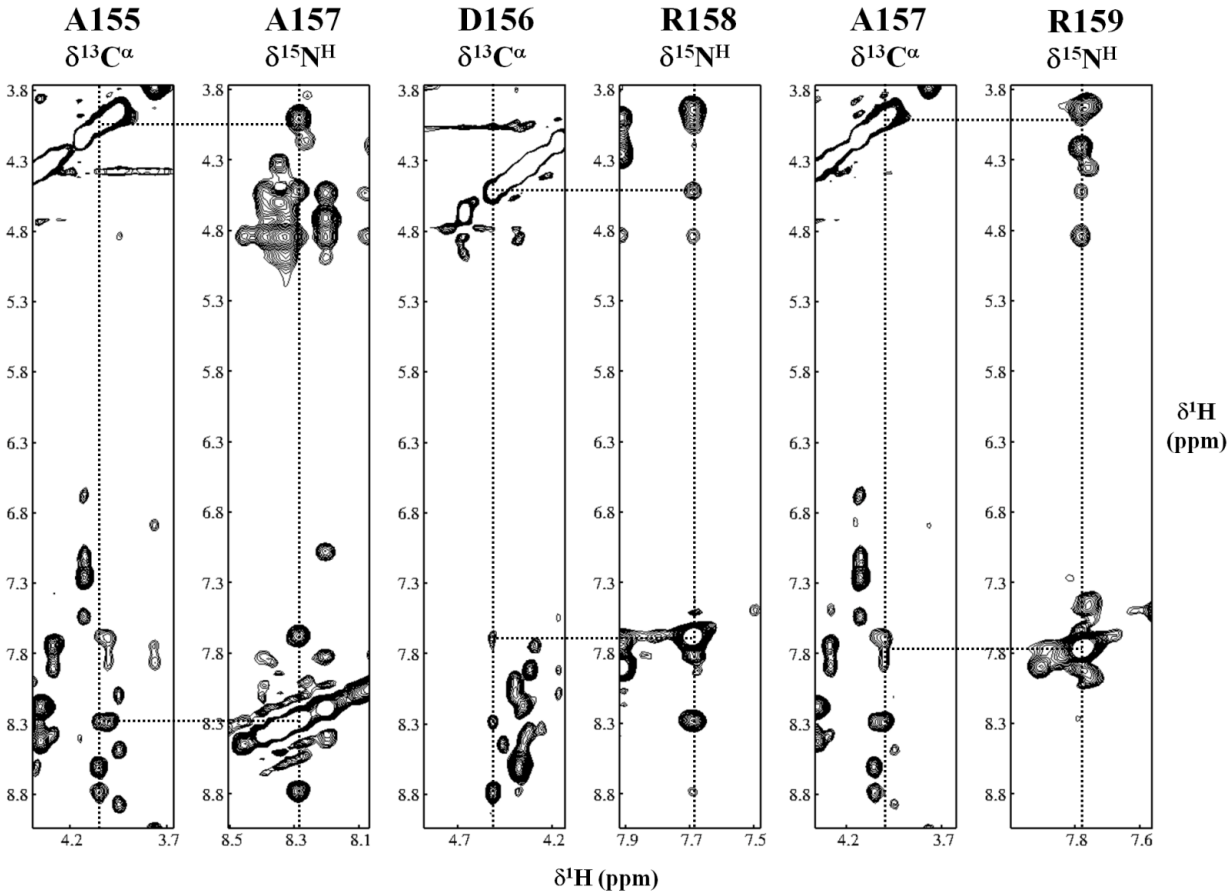

A155
δ13Cα
A157
δ15NH
D156
δ13Cα
R158
δ15NH
A157
δ13Cα
R159
δ15NH
δ1H (ppm)
δ1H (ppm)

Fig. 16. NOE correlations of 3_{10}-helix region from η1 in the BldD-CTD structure. Above vertical dotted lines in each spectrum, corresponding residue and ^{1}H chemical shift are labeled. Correlations of $d_{\alpha N}(i,i+2)$ between spectra are represented horizontal dotted lines.

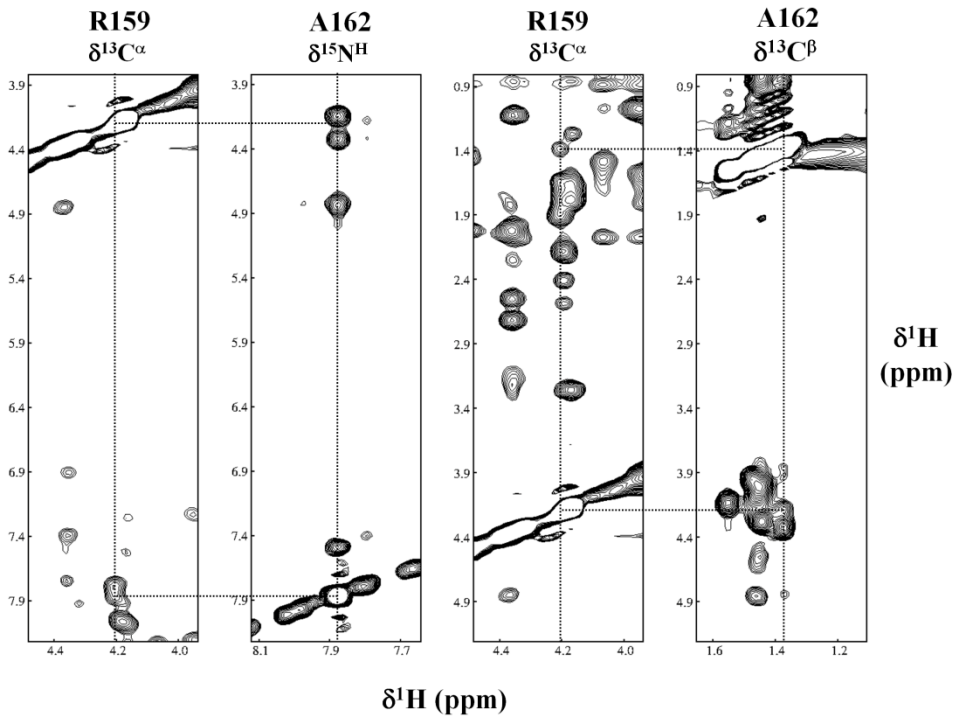


Fig. 17. NOE correlations of α-helix region from η1 in the BldD-CTD structure. Above vertical dotted lines in each spectrum, corresponding residue and ^{1}H chemical shift are labeled. Correlations of $d_{\alpha N}(i,i+3)$ or $d_{\alpha\beta}(i,i+3)$ between spectra are represented horizontal dotted lines.

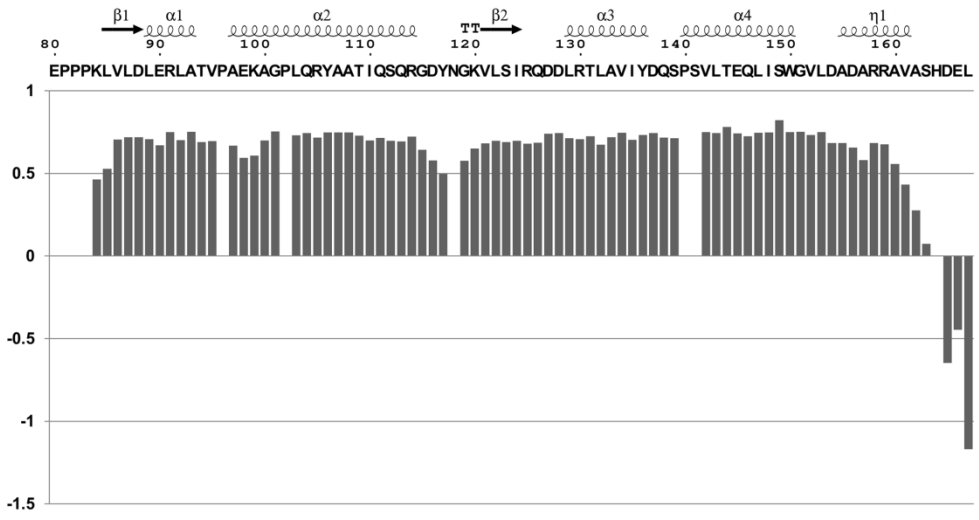


Fig. 18. [^{1}H,^{15}N]-heteronuclear NOE values of individual residues in the BldD-CTD. Secondary structure elements are illustrated over the BldD-CTD sequence, of which every tenth residue number is indicated. 'TT' represents a strict β-turn.

a structural scaffold that anchor α2 and α3. Helix α3 of BldD-NTD functions as a ‘recoginition helix’ (Kim *et al*., 2007). Likewise, BldD-CTD adopts a fold which is maintained through hydrophobic interaction of four α-helices. In the BldD-CTD structure, differently from BldD-NTD, a pair of β-sheet forming from β1 and β2 also contributes to stabilization of overall structure by which is packed against bundle of four α-helices. Interestingly, helix α3 and α4 shows a similar fold with the helix-turn-helix fold of BldD-NTD, which are anchored by helix α2 like the helix α1 of BldD-NTD (Fig. 19). However, electrostatic surface potential of BldD-CTD shows relatively negative charges (Fig. 21) while that of BldD-NTD represents strong positive charges (Fig. 20). Therefore, it might be reasonable that BldD-CTD has another function which is not related to DNA-binding.

10. Structural homology of BldD-CTD

The 3D-structure of the BldD-CTD was applied to DALI server (Holm *et al*., 2010) for searching structural homologues of the BldD-CTD. Unfortunately, searching the DALI server with BldD-CTD structure did not yield a structural homologue with a remarkable resemblance (Z-score > 4.0). However, most of the structures with relatively high Z-scores and low RMSD

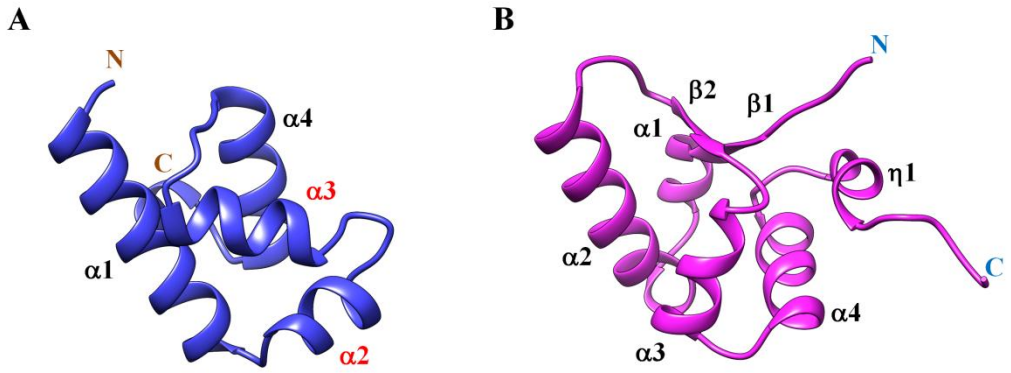


Fig. 19. Cartoon representation of (A) BldD-NTD (residues 1-79) and (B) BldD-CTD (residues 80-167). Secondary structure elements are labeled. 'N' and 'C' represent N-terminus and C-terminus, respectively. In panel (A), α2 and α3 constitutes a helix-turn-helix fold of BldD-NTD, designated as red fonts.

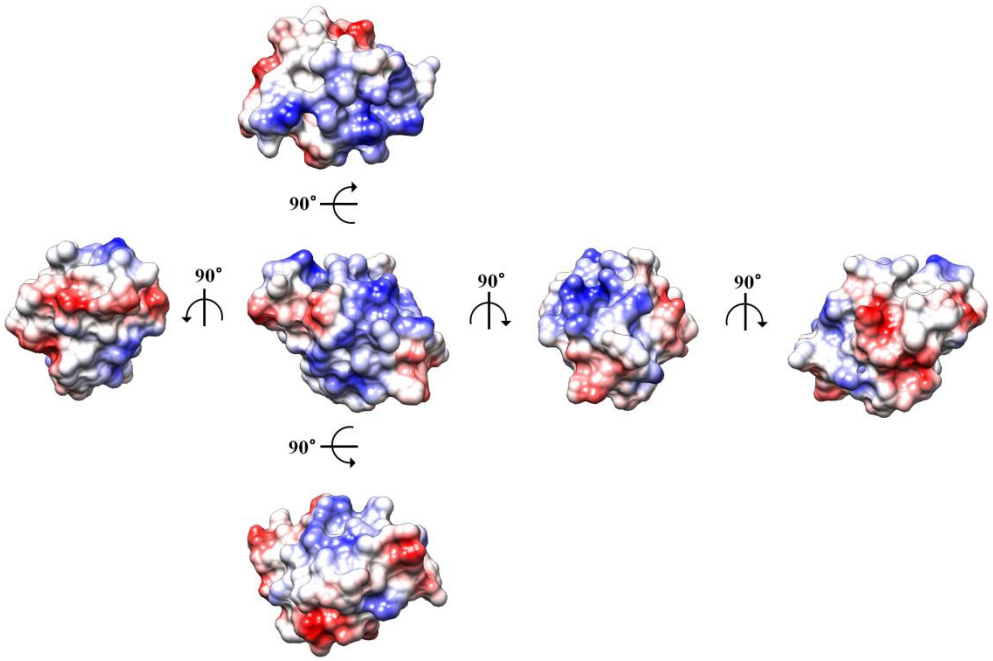


Fig. 20. Electrostatic surface potential of the BldD-NTD structure. Rendered surfaces are colored according to electrostatic surface potential at ±10 KT/e for positive (blue), negative (red) or neutral (white) charge potential. Angles and directions for rotation are labeled.

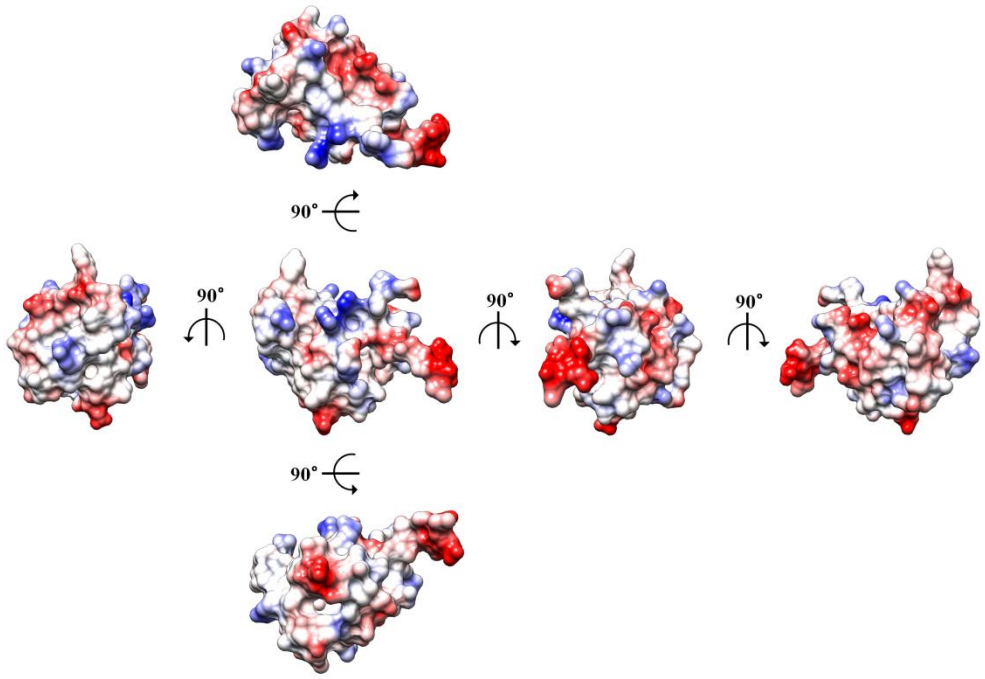


Fig. 21. Electrostatic surface potential of the BldD-CTD structure. Rendered surfaces are colored according to electrostatic surface potential at ±10 KT/e for positive (blue), negative (red) or neutral (white) charge potential. Angles and directions for rotation are labeled.

values were DNA-binding proteins containing a winged helix-turn-helix (wHTH) motif. Among them, the DNA-binding domains (DBDs) of the *Streptococcus gordonii* ScaR (PDB ID 3HRS; Z-score 2.4) and the *Thermotoga maritima* LexA (PDB ID 3K2Z; Z-score 2.6) proteins showed the most similar fold to the BldD-CTD, with an RMSD of 1.5 Å over 39 matched C^{α} atoms (Fig. 22).

ScaR-DBD and LexA-DBD (Stoll *et al*., 2009; Butala *et al*., 2009) both show a canonical winged-helix domain (WHD) with a topology of αααββ (Fig. 23). In spite of the different topology, the wHTH-like overall fold of the BldD-CTD superimposes well with those of the ScaR-DBD and LexA-DBD, particularly in the α3-α4 region (from L129 to W150), which corresponds to the helix-turn-helix (α2-α3) of the ScaR-DBD (RMSD 0.917 Å over 22 C^{α} atoms) and LexA-DBD (RMSD 1.068 Å over 22 C^{α} atoms) as shown in Fig. 24. In addition, the short two-stranded anti-parallel β-sheet (β1 and β2) of the BldD-CTD spatially compensates for the typical β-hairpin wing of the ScaR-DBD and LexA-DBD (Fig. 25).

The WHDs can interact with DNA targets thorough their wing and HTH. The wing which is the loop connecting two β-strands contacts to the minor groove of DNA and the helix-turn-helix which is formed by helix α2

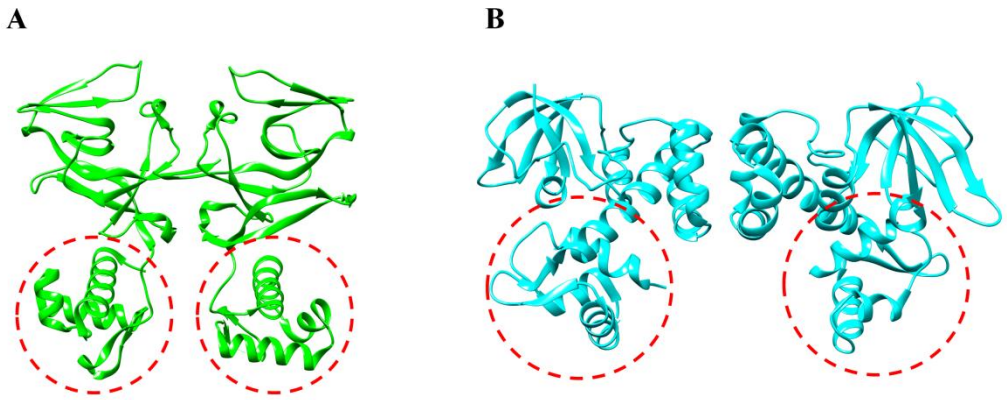


Fig. 22. Overall structures of homologues structurally similar to the BldD-CTD structure. (A) LexA from *Thermotoga maritima* (PDB ID 3K2Z) and (B) ScaR from *Streptococcus gordonii* (PDB ID 3HRS). The folds structurally similar to the BldD-CTD structure are highlighted by red-dashed circle.

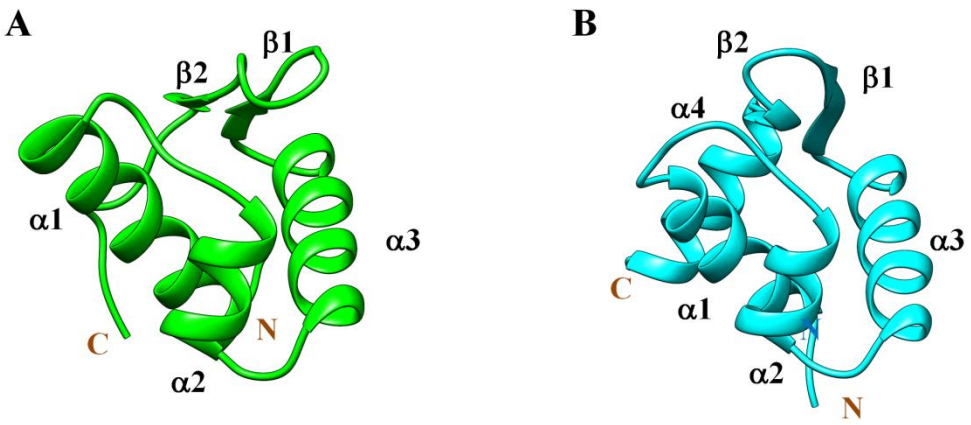


Fig. 23. Winged-helix domain of LexA (A) and ScaR (B). Secondary structure elements are labeled. ‘N’ and ‘C’ represent N-terminus and C-terminus, respectively.

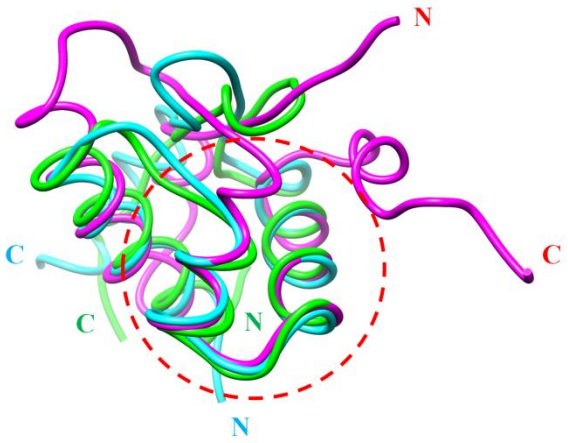


Fig. 24. Licorice style representation of superimposed BldD-CTD (magenta), LexA (green), and ScaR (cyan). The C^{α} atoms in the regions A97-Q111 and D128-G151of the BldD-CTD, E6-N20 and T23-L46 of the ScaR-WHD, and R9-G23 and S27-E50 of the LexA-WHD were superimposed. The α3-α4 region of the BldD-CTD overlays well with the corresponding regions in the other structures (red-dashed circle).

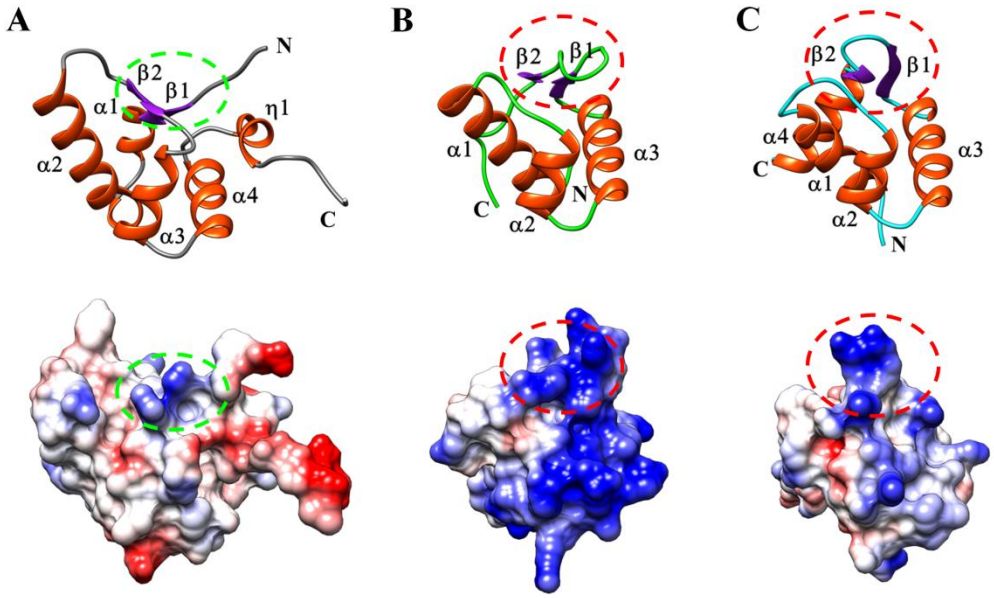


Fig. 25. Cartoon representation and electrostatic surface potential of BldD-CTD (A), LexA (B), and ScaR (C). (Top panel) Secondary structure elements are labeled. 'N' and 'C' represent N-terminus and C-terminus, respectively. (Bottom panel) Rendered surfaces are colored according to electrostatic surface potential at ±10 KT/e for positive (blue), negative (red) or neutral (white) charge potential. The β-sheet structure of BldD-CTD (A) is highlighted by green-dashed circles, and the β-hairpin structures (wing structures) of LexA and ScaR are red-dashed circles.

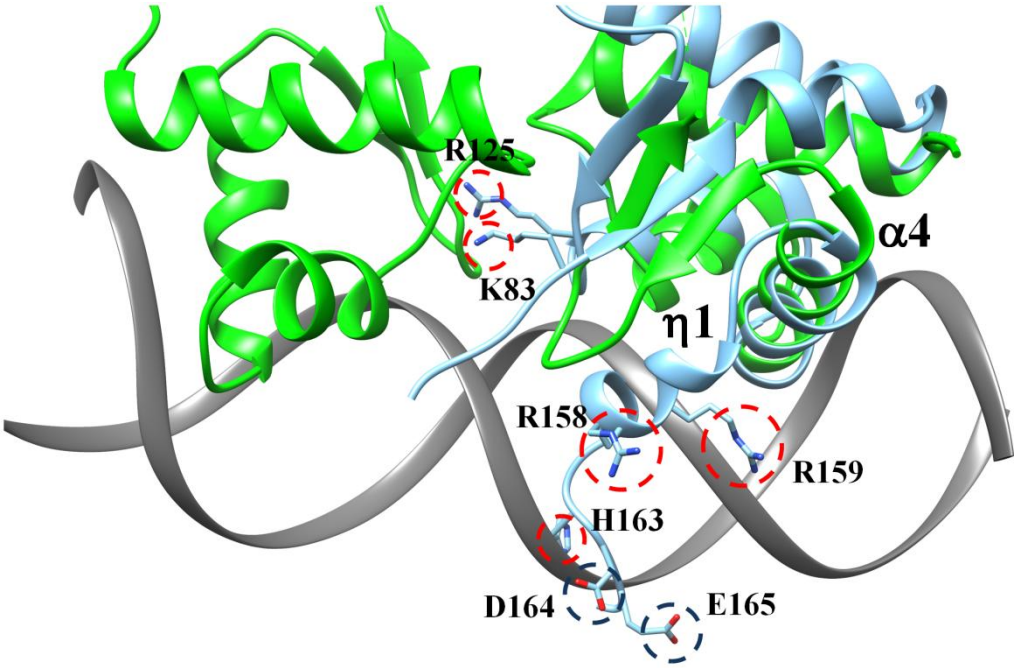


Fig. 26. Cartoon representation of superimposed BldD-CTD and LexA-DNA complex. BldD-CTD and LexA are colored cyan and green, respectively. DNA is colored gray. Positive-charged residues from wing-like structure and helix η1 from BldD-CTD are exhibited and highlighted by red-dashed circle. Negative-charged residues from C-terminus of BldD-CTD are also exhibited and highlighted by blue-dashed circle. Heteroatoms are colored red and blue for oxygen and nitrogen, respectively.

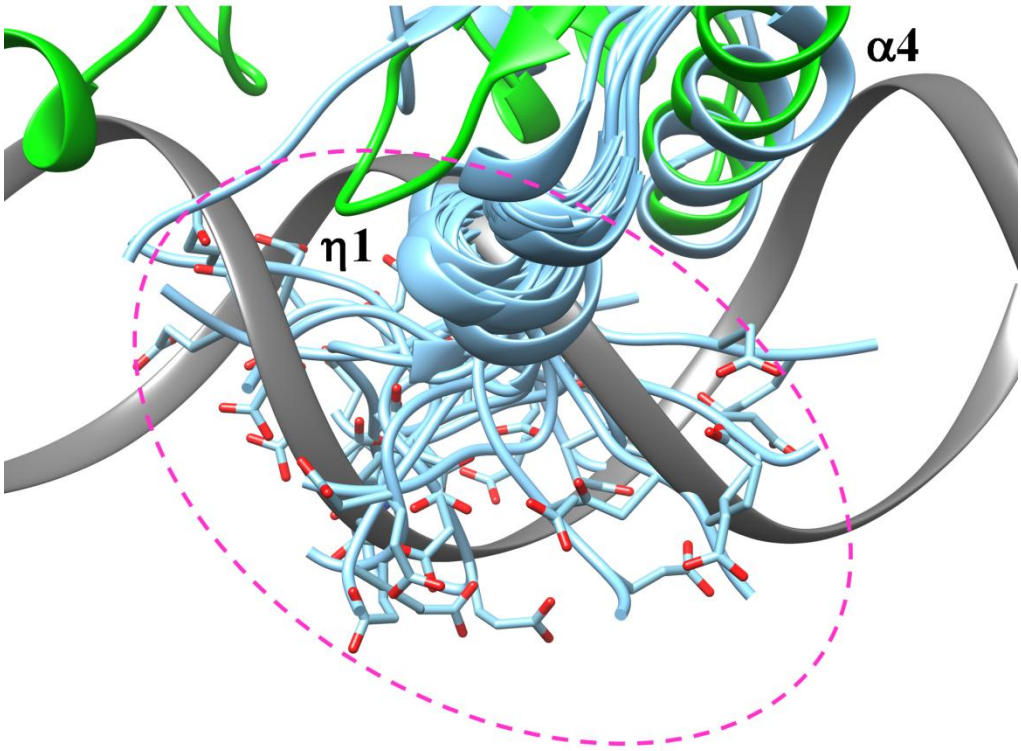


Fig. 27. Dynamic property of the BldD-CTD helix η1 and C-terminus from superimposed BldD-CTD and LexA-DNA complex. BldD-CTD and LexA are colored cyan and green, respectively. DNA is colored gray. Ensemble structure of BldD-CTD from V150 to L167 is only shown. Side-chains for D165 and E166 are shown. Heteroatom oxygen is colored red. Dynamic property of D165 and E166 is indicated by magenta-dashed line.

and α3 binds the major groove of the DNA with its helix α3 as a recognition helix (Aravind *et al*., 2005; Harami *et al*., 2013). The electrostatic surface potential analysis of the BldD-CTD shows that the face of HTH region has slightly negative charges, especially focused on its C-terminus (Fig. 27A), while its structural homologues, ScaR-DBD and LexA-DBD, have strong positive-charged surfaces on the faces of helix-turn-helix which are willing to bind the phosphate backbone of DNA (Fig. 27B and 27C). In superimposed structure of the BldD-CTD to LexA-DNA complex, hybrid helix η1 which is not existed in typical winged helix domains shows collision with DNA and negative charge residues on C-terminus of BldD-CTD (D165 and E166) might disturb DNA-binding of the BldD-CTD, although several positive charge might be positioned adjacent to DNA (Fig. 26). In addition, dynamic property of helix η1 and C-terminus could disturb the interaction between DNA and BldD-CTD (Fig. 27). Therefore, although the BldD-CTD structure is highly similar to WHD structure, it can be suggested that intact BldD-CTD has no DNA-binding activity.

11. Identification of truncated BldD

To further investigate the roles of BldD-CTD related to regulatory mechanism of BldD, the immunoblot analysis of BldD was performed

depending on morphological development of *S. coelicolor*. Interestingly, synchronized with formation of aerial mycelium, BldD appeared to be truncated in *S. coelicolor* (Fig. 28). To recognize whether BldD was truncated in N- or C-terminal region, pSET162 vector containing coding sequence of N-terminally FLAG-tagged BldD was complemented to *ΔbldD* strain. Although it could not fully complement the bald phenotype of *ΔbldD* possibly due to disturbance of attached FLAG-tag to action of BldD, delayed morphological development of this mutant strain when compared to wild type allowed to detect N-terminally FLAG-tagged BldD and its truncated form (Fig. 29A and 29B). As a result, it was confirmed that BldD was truncated in its C-terminal region.

To characterize how residues were truncated in the C-terminal region of BldD, the immunoblot analysis of 48h-cultured *S. coelicolor* was performed with various truncated forms of BldD which were heteroexpressed from pET-3a vector system in *E. coli* (Fig. 29C). It was difficult that precise site of truncation was identified, since the controls were heteroexpressed in *E. coli* and the size resolution of SDS-PAGE was limited. However, interestingly, it was observed that the truncated form corresponding to residues 1-159 of BldD had a similar size to that of truncated BldD detected in *S. coelicolor*. This truncation of BldD indicated that C-terminal region including a portion of hybrid helix η1 (A155-A161) shown in the structure of BldD-CTD (Fig. 13) was removed when

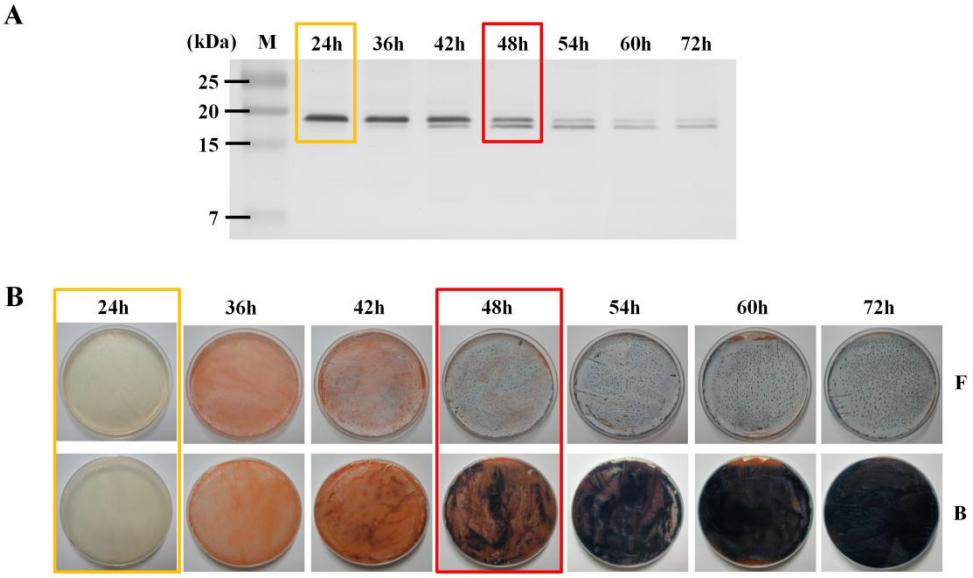


Fig. 28. Developmental stage-dependent protein patterns of BldD. (A) Immunoblot analysis of BldD in wild type. 40 μg of crude extract from each mycelia harvested at the labeled time were loaded on each lane. Cells were grown on R2YE overlaid with cellophane disc at 30 °C. (B) Phenotype of *S. coelicolor* with time-dependent morphological change. 'F' and 'B' represent front side and back side of plates, respectively.

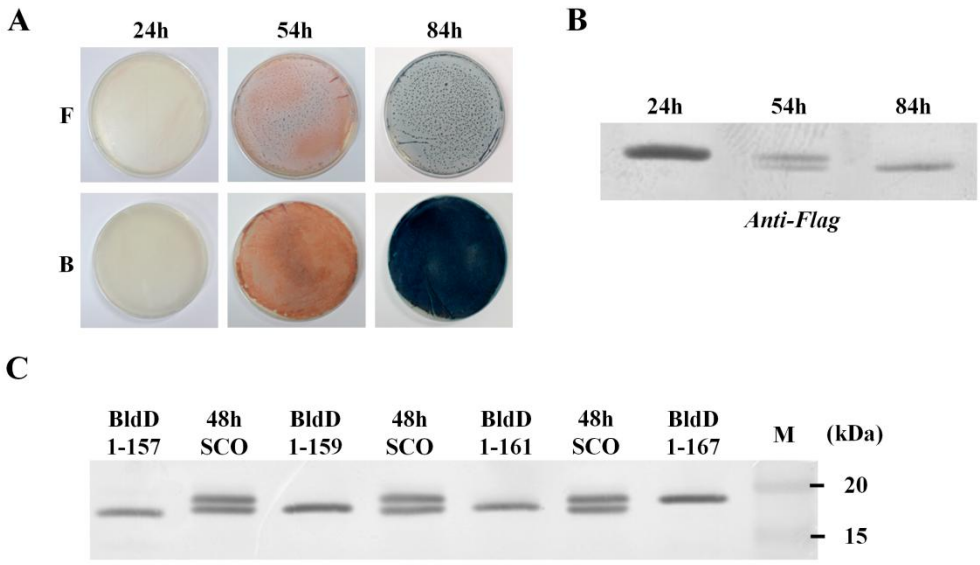


Fig. 29. Characterization of truncated BldD. (A) Phenotype of *ΔbldD* strain complemented by introduction of pSET162 derivatives containing sequence for N-terminally FLAG-tagged BldD. (B) Immunoblot analysis of N-terminally FLAG-tagged BldD. 40 μg of crude extract from each mycelia harvested at the labeled time were loaded on each lane. Cells were grown on R2YE overlaid with cellophane disc at 30 °C. (C) Verification of truncated BldD. 40 μg of crude extract from wild type cultured for 48 h were applied to immunoblot analysis. Various truncated forms of BldD heteroexpressed using pET-3a system in *E. coli* were used as standards.

S. coelicolor formed aerial mycelium.

12. Gel mobility shift assays and *in vivo* complementation experiments of truncated forms of BldD

In the stage of aerial mycelium formation of *S. coelicolor*, transcriptional level of *bldD* was considerably decreased but those of BldD-regulated genes including *bldN* and *whiG* were increased (den Hengst *et al*., 2010; Elliot *et al*., 1998; Elliot *et al*., 2001). In addition, after morphological differentiation, truncated BldD existed as small amount compared to that of substrate mycelium. These suggest that truncated BldD might act as a repressor only for regulatory region of *bldD*.

As mentioned above, in the BldD-CTD structure, the hybrid helix η1 was expected that disturb the DNA-binding of BldD-CTD (Fig. 26 and 27). In spite of negative-charged surface of BldD-CTD (Fig. 25A), it shows considerably similar structure to those of WHDs and is linked to the BldD-NTD, a strong DNA-binding domain. Thus, the truncation of helix η1 might induce BldD-CTD to bind the regulatory region of *bldD*, followed by stronger repression of truncated BldD for *bldD*. To investigate this possibility, four truncated forms of BldD-CTD and BldD were heteroexpressed in *E. coli* and purified (Fig. 30). Purified proteins were applied to gel mobility shift assays

with target DNAs of BldD. However, truncated forms of BldD-CTD did not bind any DNAs in spite of high concentration up to 100 μM (Fig. 31). On the contrary, the truncated forms of BldD showed slightly increased affinity for all DNAs (Fig. 32). As detection of truncated BldD was synchronized with morphological change of *S. coelicolor* and DNA-binding patterns were not altered in all forms of BldD, it seemed that increased DNA-binding seen in Fig. 32 was not by DNA-binding of BldD-CTD but by net charge increase due to removal of negative-charged residues, D165 and E166. In addition, *in vivo* complementation experiments for truncated forms of BldD did not show any significant difference with complementation of wild type BldD (Fig. 33). Thus, despite the unique structural features of helix η1, the role of BldD-CTD would not be related to this region.

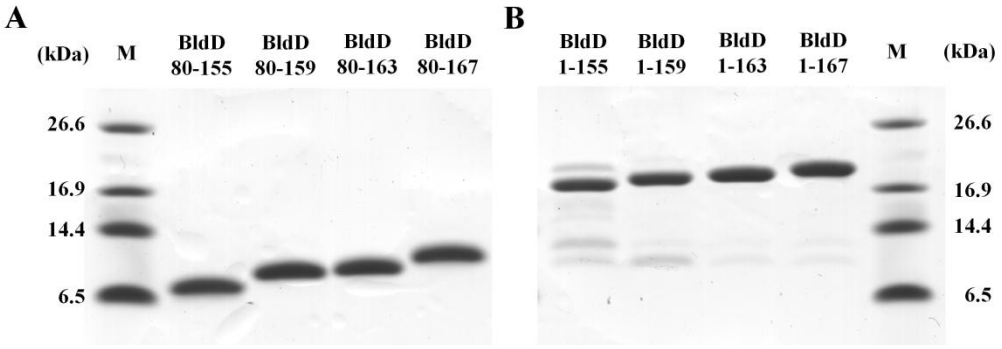


Fig. 30. Purification of various truncated forms of BldD-CTD (A) and BldD (B). 2 μg of each purified protein were loaded on 15.4 % Tris-tricine SDS-PAGE.

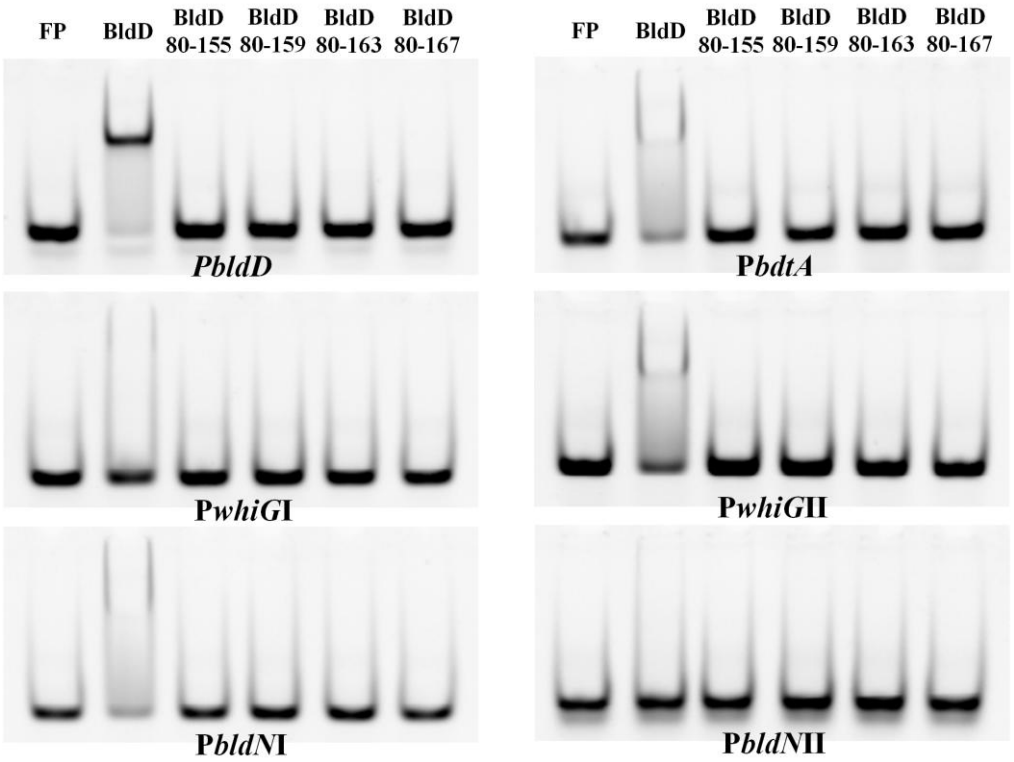


Fig. 31. Gel mobility shift assay of various truncated forms of BldD-CTD against six BldD-binding sites. 100 μM of each truncated form of BldD-CTD were mixed with each probe and loaded onto 8 % native gel containing 1X TBE. 5 μM of BldD were used as a control. Each gel was visualized with 60s exposure time in SYBR mode using LAS3000 (Fujifilm).

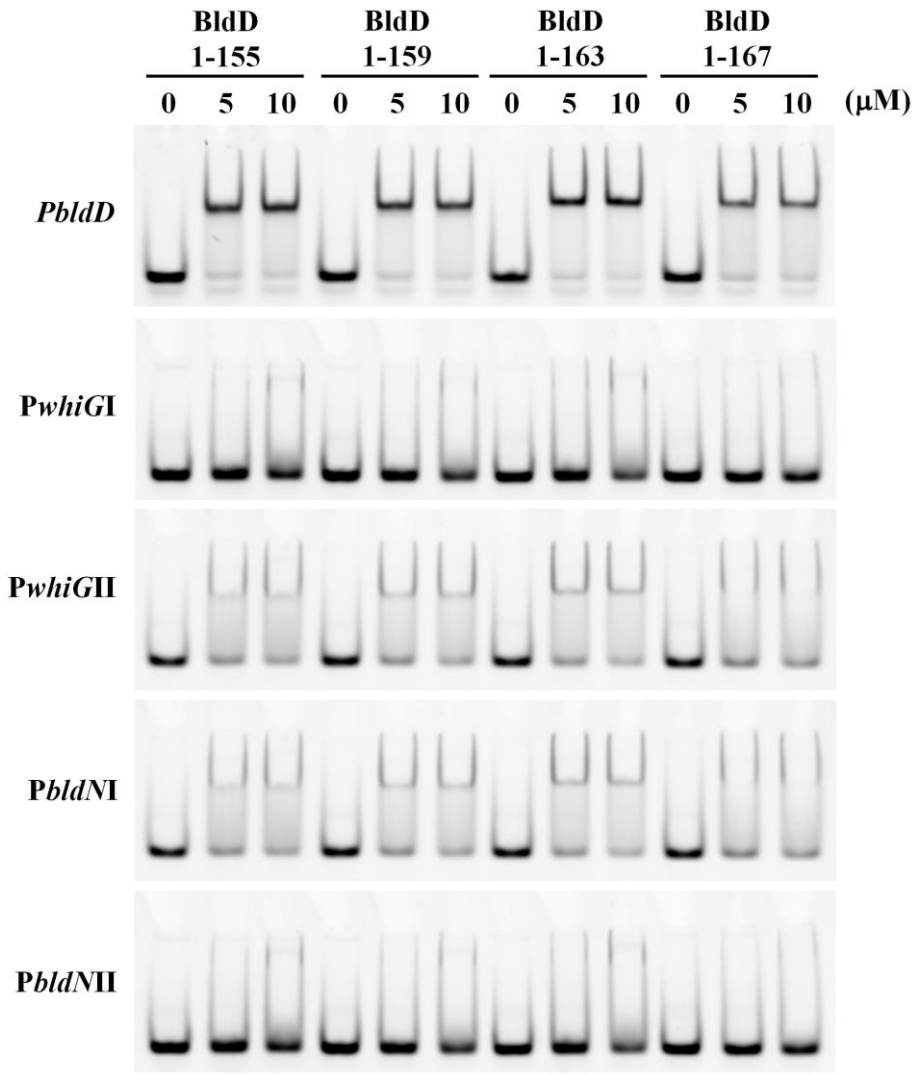


Fig. 32. Gel mobility shift assay of various truncated forms of BldD-CTD against six BldD-binding sites. 5 μM or 10 μM of each protein were mixed with each probe and loaded onto 8 % native gel containing 1X TBE. Each gel was visualized with 60s exposure time in SYBR mode using LAS3000 (Fujifilm).

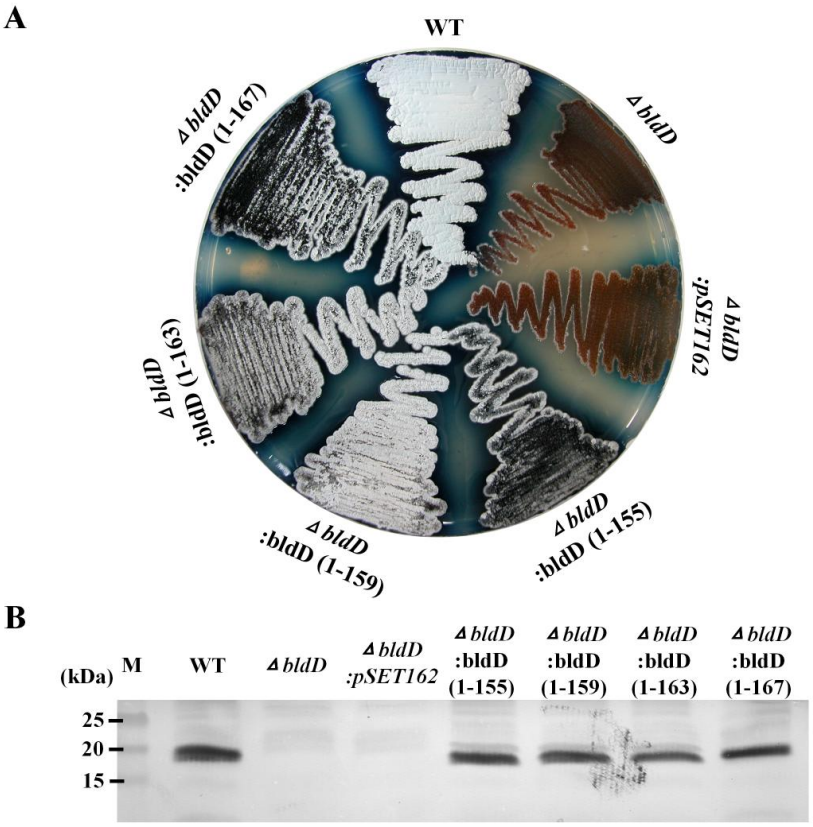


Fig. 33. *In vivo* complementation experiments of BldD truncated forms. (A) Phenotype of wild type and *ΔbldD* starains complemented by introduction of pSET162 derivatives containing wild type (WT) *bldD* and truncated *bldD*. pSET162 was used as a negative control. All strains were grown on R2YE medium for 4 days at 30 °C. (B) Immnuoblot analysis of wild type BldD and fragmented BldD. 40 μg of crude extract from each strain were loaded on each lane. All strains were harvested after grown on R2YE overlaid with cellophane disc for 1 day at 30 °C.

IV. Discussion

The solution structure of BldD-CTD has been determined and has a similar fold with winged-helix domain (WHD). The structure of the BldD-CTD contains an additional C-terminal helix, η1, which appears to be dynamic and might therefore interfere with DNA-binding of BldD-CTD. But, C-terminal truncated forms of the BldD-CTD, wherein the η1 was truncated, were also incapable of binding DNA. Generally, strongly positive-charged surface of the helix-turn-helix face is conserved in the DNA-binding WHDs and favors the interaction with the negatively charged phosphate backbone of DNA (Aravind *et al*., 2005; Harami *et al*., 2013). In contrast, the electrostatic surface potential of the BldD-CTD shows abundant negative charges rather than positive charges. Thus, anionic property of the BldD-CTD appears to be mainly responsible for its inability to bind DNA.

Although the major function of WHD is DNA-binding, WHD have variable structures and functions beyond DNA-binding. The CTD of RAP74 (the large subunit of transcription factor IIF) adopts a WHD fold and interacts with the CTD of FCP1 (RNA polymerase II C-terminal domain phosphatase) through conserved hydrophobic residues exposed from α2 and α3 (Nguyen *et al*., 2003) (Fig. 34A, 35A and 35B). Likewise, the WHD of DP2 forms

heterodimeric transcription factor with E2F4 through hydrophobic patch exposed from α1 and α3 (Zheng et al., 1999) (Fig. 34B, 35C and 35D). These suggest an alternative role for the BldD-CTD as a protein-protein interaction module. As shown in Fig. 36, the region from R105 to D128 of the BldD-CTD is strictly conserved among its homologues. In addition, ConSurf analysis (Glaser et al., 2003) also indicated highly conservation of this region (Fig. 37). Interestingly, like RAP74 and DP2, a hydrophobic surface patch is formed by the highly conserved Y106, I110, Y117, V121, L122, I124, and I135 in the BldD-CTD structure (Fig. 38). Therefore, the conserved hydrophobic surface patch on the BldD-CTD could provide a suitable interface for putative PPIs. In addition, the hydrophobic surface patch is surrounded by highly conserved charged residues, such as R105, R114, R125, D116 and D128 (Fig. 38). This could facilitate the anchoring of inbound proteins through salt-bridge formation, similar to interactions shown in the RAP74-FCP1 complex and DP2-E2F4 heterodimer (Nguyen *et al*., 2003; Zheng et al., 1999).

In conclusion, the solution structure of the BldD-CTD reveals a novel type of WHD fold that is suitable for protein-protein interactions, rather than DNA binding. Therefore, it can be suggested a possible role of the BldD-CTD as a molecular adaptor, switching transcriptional regulation through a protein-protein interaction.

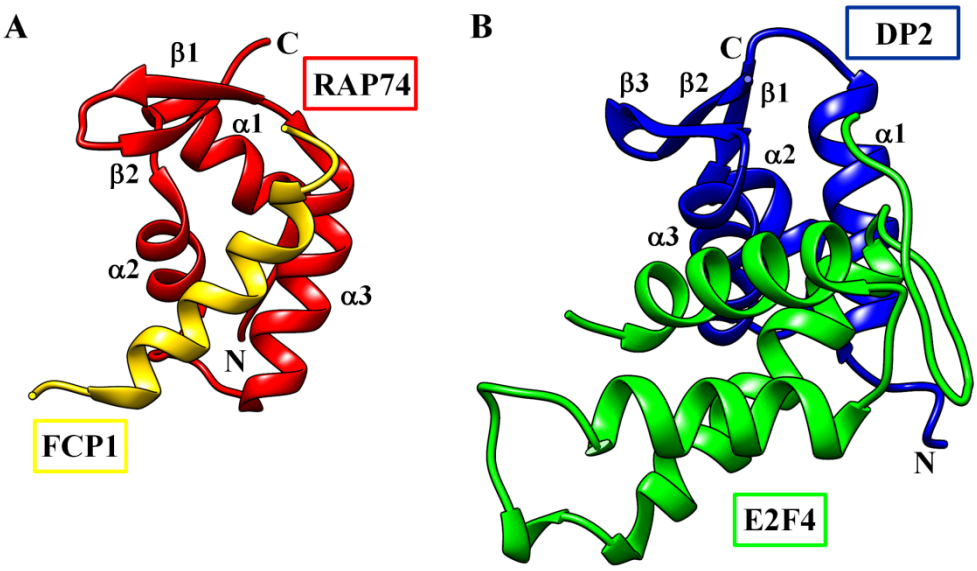


Fig. 34. Cartoon representation of RAP74-FCP1 complex and DP2-E2F4 heterodimer. (A) RAP74-FCP1 complex structure (Nguyen *et al*., 2003). RAP74-CTD and FCP1-CTD are colored red and yellow, respectively. Secondary structure elements of RAP74-CTD are labeled only. (B) DP2-E2F4 heterodimer structure (Zheng *et al*., 1999). DP2 and E2F4 are colored blue and green, respectively. Secondary structure elements of DP2 are labeled only.

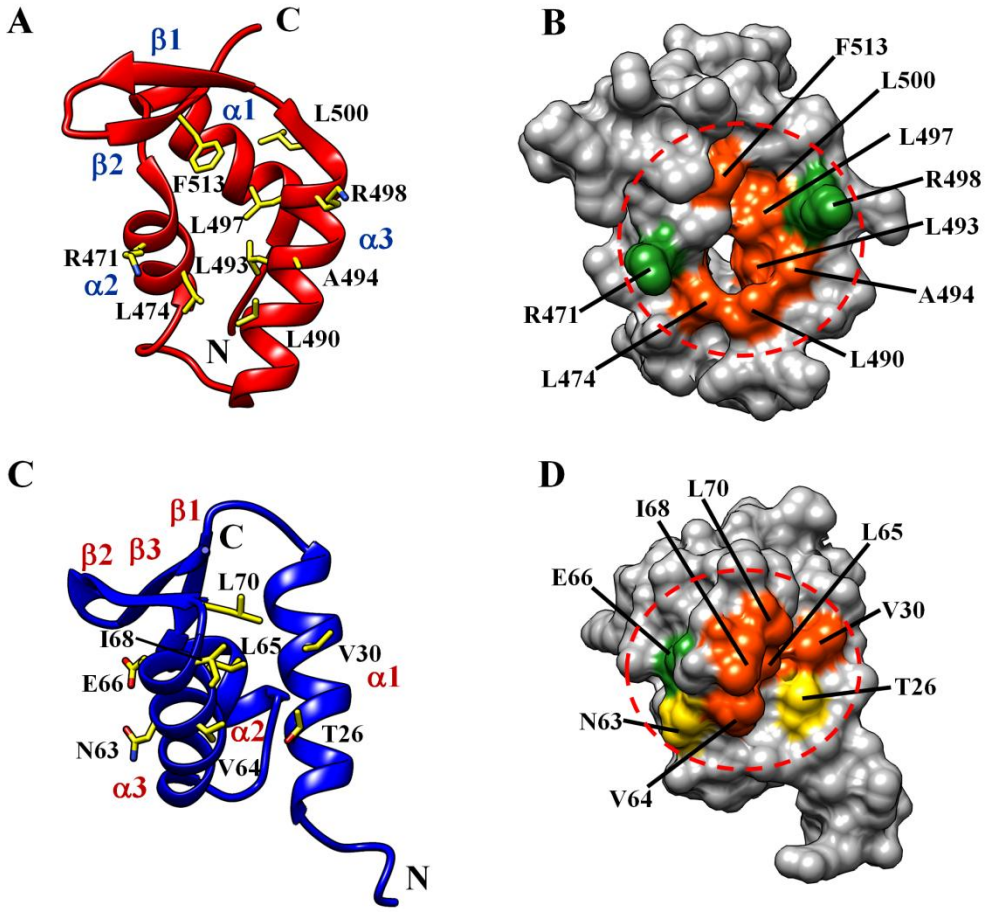


Fig. 35. Hydrophobic patch in winged-helix domain of RAP74 and DP2. Cartoon representation showing hydrophobic patch in the RAP74-CTD (A) and DP2 (C). Surface representation of RAP74-CTD (B) and DP2 (D) showing highly conserved residues. Secondary structure elements are labeled. 'N' and 'C' represent N-terminus and C-terminus, respectively. In panel (B) and (D), hydrophobic, neutral and electrically charged residues are labeled and colored orange, yellow and green, respectively.

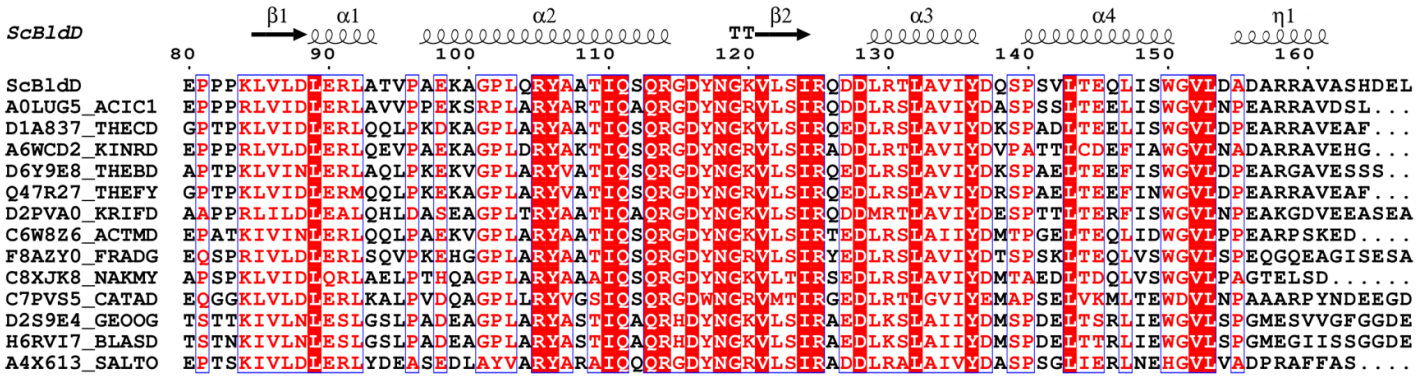

ScBldD
β1
α1
α2
TT
β2
α3
α4
η1
80
90
100
110
120
130
140
150
160
ScBldD EPPPKLVLDLERLATVPAEKAGPLQRYAATIQSQRGDYNGKVLSIRQDDLRTLAVIYDQSPSVLTEQLISWGVLDADARRAVASHDEL
A0LUG5_ACIC1 EPPPRLVIDLERLAVVPPEKSRPLARYARTIQAQRGDYNGRVLSIRQDDLRSLAVIYDASPSSLTEELISWGVLNPEARRAVDSL...
D1A837_THECD GPTPKLVIDLERLQQLPKDKAGPLARYAATIQSQRGDYNGKVLSIRQEDLRSLAVIYDKSPADLTEELISWGVLDPEARRAVEAF...
A6WCD2_KINRD EPPPRLVLDLERLQEVPAEKAGPLDRYAKTIQSQRGDYNGRVLSIRADDLRTLAVIYDVPATTLCDEFIAWGVLNADARRAVEHG...
D6Y9E8_THEBD APTPKLVINLERLAQLPKEKVGPLARYVATIQSQRGDYNGKVLSIRQEDLRSLAVIYDKSPAELTEEFISWGVLDPEARGAVESSS..
Q47R27_THEFY GPTPKLVIDLERMQQLPKEKAGPLARYVATIQSQRGDYNGRVLSIRQEDLRSLAVIYDRSPAELTEEFINWGVLDPEARRAVEAF...
D2PVA0_KRIFD AAPPRLILDLEALQHLDASEAGPLTRYAATIQAQRGDYNGKVLSIRQDDMRTLAVIYDESPTTLTERFISWGVLNPEAKGDVEEASEA
C6W8Z6_ACTMD EPATKIVINLERLQQLPAEKVGPLARYAATIQSQRGDYNGKVLSIRTEDLRSLAIIYDMTPGELTEQLIDWGVLPPEARPSKED....
F8AZY0_FRADG EQSPRIVLDLERLSQVPKEHGGPLARYAATIQSQRGDYNGRVLSIRYEDLRSLAVIYDTSPSKLTEQLVSWGVLSPEQGQEAGISESA
C8XJK8_NAKMY APSPKLVIDLQRLAELPTHQAGPLARYAAAIQSQRGDYNGKVLTIRSEDLRSLAVIYDMTAEDLTDQLVSWGVLPAGTELSD......
C7PVS5_CATAD EQGGKLVLDLERLKALPVDQAGPLLRYVGSIQSQRGDWNGRVMTIRGEDLRTLGVIYEMAPSELVKMLTEWDVLNPAAARPYNDEEGD
D2S9E4_GEOOG TSTTKIVLNLESLGSLPADEAGPLARYASTIQAQRHDYNGKVLSIRAEDLKSLAIIYDMSPDELTSRLIEWGVLSPGMESVVGFGGDE
H6RVI7_BLASD TSTNKIVLNLESLGSLPADEAGPLARYASTIQAQRHDYNGKVLSIRAEDLKSLAIIYDMSPDELTTRLIEWGVLSPGMEGIISSGGDE
A4X613_SALTO EPTSKIVLDLERLYDEASEDLAYVARYARAIQQQRGDYNGRVLSIRADDLRALAIVYDASPSGLIERLNEHGVLVADPRAFFAS....

Fig. 36. Multiple sequence alignment of the BldD-CTD with minimized redundancy. Multiple sequence alignment of the BldD-CTD and its representative homologues after removing redundancy, analyzed by ClustalW and displayed using ESPript. Secondary structure elements are illustrated over the *Streptomyces coelicolor* (Sc) BldD-CTD sequence, of which every tenth residue number is indicated. 'TT' represents a strict β-turn. Each of the other protein sequences is labeled on the left with the UniProt (http://www.uniprot.org) accession number, followed by an abbreviated name of the organism (ACIC1, *Acidothermus cellulolyticus*; THECD, *Thermomonospora curvata*; KINRD, *Kineococcus radiotolerans*; THEBD, *Thermobispora bispora*; THEFY, *Thermobifida fusca*; KRIFD, *Kribbella flavida*; ACTMD, *Actinosynnema mirum*; FRADG, *Frankia symbiont* subsp. *Datisca glomerata*; NAKMY, *Nakamurella multipartita*; CATAD, *Catenulispora acidiphila*; GEOOG, *Geodermatophilus obscurus*; BLASD, *Blastococcus saxobsidens*; SALTO, *Salinispora tropica*). The fully conserved amino acids are highlighted with red-shaded boxes, while the residues with conservative substitutions are colored red and blue-boxed.

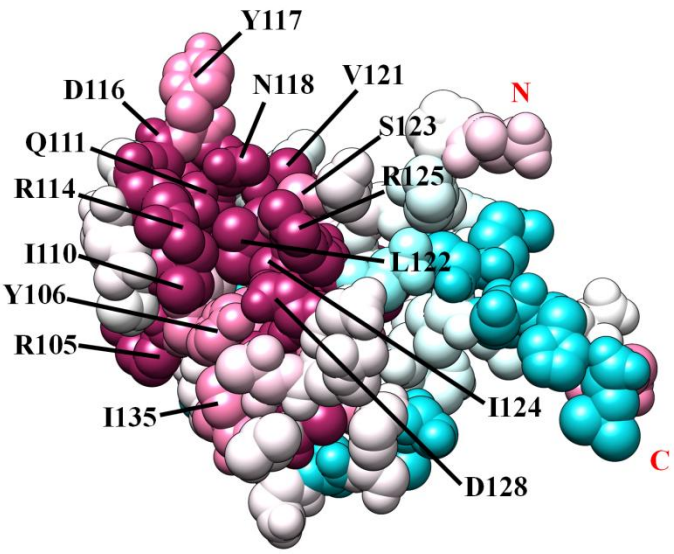


Fig. 37. Conserved surface coloring of the BldD-CTD structure. Consurf image of BldD-CTD representing the conserved residues from the multiple sequence alignment of 34 BldD-CTD homologues using ClustalW. The degree of residue conservation is colored with ranges from magenta (highly conserved) to cyan (variable). Highly conserved residues are labeled. ‘N’ and ‘C’ represent N-terminus and C-terminus, respectively.

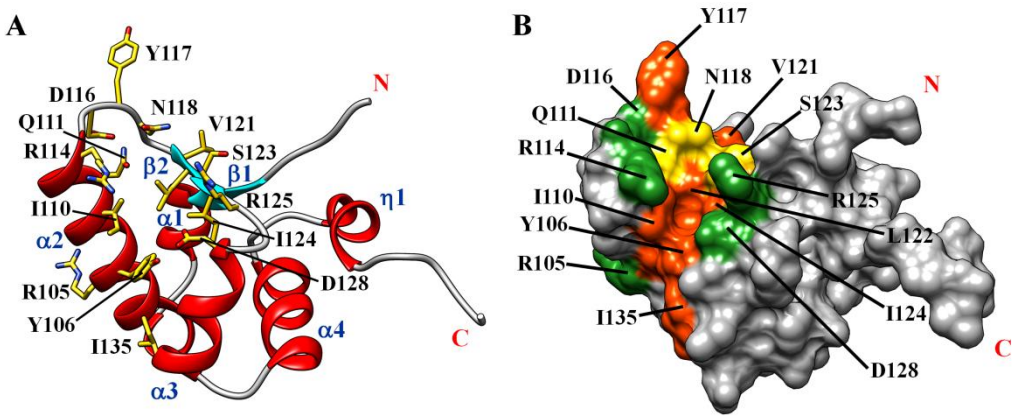


Fig. 38. Conserved hydrophobic patch of the BldD-CTD. (A) Cartoon representation showing conserved hydrophobic patch in the BldD-CTD structure. Side-chains are represented as yellow sticks with heteroatoms colored red for oxygen and blue for nitrogen. Secondary structure elements are colored red for helices (α1-α4 and η1) and cyan for β-strands (β1 and β2). (B) Surface representation of BldD-CTD showing highly conserved residues. Hydrophobic, neutral and electrically charged residues are colored orange, yellow and green, respectively. 'N' and 'C' represent N-terminus and C-terminus, respectively.

V. REFERENCES

den Hengst, C.D., Tran, N.T., Bibb, M.J., Chandra, G., Leskiw, B.K., Buttner, M.J. (2010) Genes essential for morphological development and antibiotic production in *Streptomyces coelicolor* are targets of BldD during vegetative growth. *Mol Microbiol* **78**:361-379.

Ainsa, J. A., Ryding, N. J., Hartley, N., Findlay, K. C., Bruton, C. J. & Chater, K. F. (2000). WhiA, a protein of unknown function conserved among gram-positive bacteria, is essential for sporulation in *Streptomyces coelicolor* A3(2). *J Bacteriol* **182**, 5470-5478.

Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M., Iyer, L.M. (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* **29**:231-262.

Bhattacharya, A., Tejero, R., Montelione, G.T. (2007) Evaluating protein structures determined by structural genomics consortia. *Proteins* **66**:778-795.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., Warren, G.L. (1998)

Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**:905-921.

Butala, M., Zgur-Bertok, D., Busby, S.J. (2009) The bacterial LexA transcriptional repressor. *Cell Mol Life Sci* **66**:82-93.

Champness, W. C. (1988). New loci required for *Streptomyces coelicolor* morphological and physiological differentiation. *J Bacteriol* **170**, 1168-1174.

Chater, K. F. & Merrick, M. J. (1976). Approaches to the study of differetiation in *Streptomyces coelicolor* A3(2). In *Second international symposium on the genetics of industrial microorganisms*, pp. 583-593. Edited by K. D. MacDonald. London, United Kingdom: Academic Press.

Chater, K. F. (1972). A morphological and genetic mapping study of white colony mutants of *Streptomyces coelicolor*. *J Gen Microbiol* **72**, 9-28.

Chater, K. F. (1993). Genetics of differentiation in Streptomyces. *Annu Rev Microbiol* **47**, 685-713.

Chater, K. F., Bruton, C. J., Plaskitt, K. A., Buttner, M. J., Mendez, C. and Helmann, J. D. (1989) The developmental fate of *S. coelicolor* hyphae depends upon a gene product homologous with the motility factor of *B. subtilis*. *Cell* **59 :** 133-143.

Davis, N. K. & Chater, K. F. (1992). The *Streptomyces coelicolor whiB* gene

encodes a small transcription factor-like protein dispensable for growth but essential for sporulation. *Mol Gen Genet* **232**, 351-358.

Delaglio, F., Grzesiek, S., Vuister, GW., Zhu, G., Pfeifer, J., Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**:277-293.

Elliot, M. A. & Leskiw, B. K. (1999). The BldD protein from *Streptomyces coelicolor* is a DNA-binding protein. *J Bacteriol* **181**, 6832-6835.

Elliot, M., Damji, F., Passantino, R., Chater, K. & Leskiw, B. (1998). The *bldD* gene of *Streptomyces coelicolor* A3(2): a regulatory gene involved in morphogenesis and antibiotic production. *J Bacteriol* **180**, 1549-1555.

Elliot, M. A., Bibb, M. J., Buttner, M. J. & Leskiw, B. K. (2001). BldD is a direct regulator of key developmental genes in *Streptomyces coelicolor* A3(2). *Mol Microbiol* **40**, 257-269.

Elliot, M.A., Locke, T.R., Galibois, C.M., Leskiw, B.K.(2003) BldD from *Streptomyces coelicolor* is a non-essential global regulator that binds its own promoter as a dimer. *FEMS Microbiol Lett* **225**:35-40.

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., Kay, L.E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry*

33:5984-6003.

Fernandez-Moreno, M. A., Caballero, J. L., Hopwood, D. A. & Malpartida, F. (1991). The act cluster contains regulatory and antibiotic export genes, direct targets for translational control by the *bldA* tRNA gene of Streptomyces. *Cell* **66**, 769-780.

Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., Ben-Tal, N. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **19**:163-164.

Güntert, P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**:353-378.

Güntert, P., Mumenthaler, C., Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* **273**:283-298.

Harami, G.M., Gyimesi, M., Kovács, M. (2013) From keys to bulldozers: expanding roles for winged helix domains in nucleic-acid-binding proteins. *Trends Biochem Sci* **38**:364-371.

Herrmann, T., Güntert, P., Wüthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol*

319:209-227.

Holm, L., Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**:W545-549.

Hopwood, D. A., Wildermuth, H. & Palmer, H. M. (1970). Mutants of *Streptomyces coelicolor* defective in sporulation. *J Gen Microbiol* **61**, 397-408.

Ikura, M., Kay, L.E., Bax, A. (1991) Improved three-dimensional 1H-13C-1H correlation spectroscopy of a 13C-labeled protein using constant-time evolution. *J Biomol NMR* **1**:299-304.

Johnson, B. A. (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* **278**:313-352.

Kay, L.E., Ikura, M., Tschudin, R., Bax, A. (1990) Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. *J Magn Reson* **213**:423-441.

Kelemen, G. H. & Buttner, M. J. (1998). Initiation of aerial mycelium formation in Streptomyces. *Curr Opin Microbiol* **1**, 656-662.

Kelemen, G. H., Brown, G. L., Kormanec, J., Potuckova, L., Chater, K. F. & Buttner, M. J. (1996). The positions of the sigma-factor genes, *whiG* and *sigF*, in the hierarchy controlling the development of spore chains in the aerial hyphae of *Streptomyces coelicolor* A3(2). *Mol Microbiol* **21**,

593-603.

Kelemen, G. H., Viollier, P. H., Tenor, J., Marri, L., Buttner, M. J. & Thompson, C. J. (2001). A connection between stress and development in the multicellular prokaryote *Streptomyces coelicolor* A3(2). *Mol Microbiol* **40**, 804-814.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., Hopwood, D.A. (2000) Practical *Streptomyces* Genetics. *The John Innes Foundation Norwich.*

Kim, I.K., Lee, C.J., Kim, M.K., Kim, J.M., Kim, J.H., Yim, H.S., Cha, S.S., Kang, S.O. (2006) Crystal structure of the DNA-binding domain of BldD, a central regulator of aerial mycelium formation in *Streptomyces coelicolor* A3(2). *Mol Microbiol* **60**:1179-1193.

Lawlor, E. J., Baylis, H. A. & Chater, K. F. (1987). Pleiotropic morphological and antibiotic deficiencies result from mutations in a gene encoding a tRNA-like product in *Streptomyces coelicolor* A3(2). *Genes Dev* **1**, 1305-1310.

Lee, C.J., Won, H.S., Kim, J.M., Lee, B.J., Kang, S.O. (2007) Molecular domain organization of BldD, an essential transcriptional regulator for developmental process of *Streptomyces coelicolor* A3(2). *Proteins* **68**:344-352.

Leskiw, B. K., Lawlor, E. J., Fernandez-Abalos, J. M. & Chater, K. F.

(1991). TTA codons in some genes prevent their expression in a class of developmental, antibiotic-negative, Streptomyces mutants. *Proc Natl Acad Sci U S A* **88**, 2461-2465.

Linge, J.P., Williams, M.A., Spronk, C.A., Bonvin, A.M., Nilges, M. (2003) Refinement of protein structures in explicit solvent. *Proteins* **50**:496-506.

Losick, R. & Shapiro, L. (1993). Checkpoints that couple gene expression to morphogenesis. *Science* **262**, 1227-1228.

Markley, J.L., Bax, A., Arata, Y., Hilbers, C.W., Kaptein, R., Sykes, B.D., Wright, P.E., Wüthrich, K. (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. *J Mol Biol* **280**:933-952.

Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M., Clore, G.M. (1989) Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry* **28**:6150-6156.

Merrick, M. J. (1976). A morphological and genetic mapping study of bald colony mutants of *Streptomyces coelicolor*. *J Gen Microbiol* **96**, 299-315.

Molle, V. & Buttner, M. J. (2000). Different alleles of the response regulator gene *bldM* arrest *Streptomyces coelicolor* development at distinct stages.

Mol Microbiol **36**, 1265-1278.

Molle, V., Palframan, W. J., Findlay, K. C. & Buttner, M. J. (2000). WhiD and WhiB, homologous proteins required for different stages of sporulation in *Streptomyces coelicolor* A3(2). *J Bacteriol* **182**, 1286-1295.

Nederveen, A.J., Doreleijers, J.F., Vranken, W., Miller, Z., Spronk, C.A., Nabuurs, S.B., Güntert, P., Livny, M., Markley, J.L., Nilges, M., Ulrich, E.L., Kaptein, R., Bonvin, A.M. (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* **59**:662-672.

Nepravishta, R,, Polizio, F., Paci, M., Melino, S. (2012) A metal-binding site in the RTN1-C protein: new perspectives on the physiological role of a neuronal protein. *Metallomics* 4:480-487.

Nguyen, B.D., Abbott, K.L., Potempa, K., Kobor, M.S., Archambault, J., Greenblatt, J., Legault, P., Omichinski, J.G. (2003) NMR structure of a complex containing the TFIIF subunit RAP74 and the RNA polymerase II carboxyl-terminal domain phosphatase FCP1. *Proc Natl Acad Sci USA* **100**:5688-5693.

Nodwell, J. R., McGovern, K. & Losick, R. (1996). An oligopeptide permease responsible for the import of an extracellular signal governing aerial mycelium formation in *Streptomyces coelicolor*. *Mol Microbiol* **22**, 881-

893.

Nodwell, J. R., Yang, M., Kuo, D. & Losick, R. (1999). Extracellular complementation and the identification of additional genes involved in aerial mycelium formation in *Streptomyces coelicolor*. *Genetics* **151**, 569-584.

Olejniczak, E.T., Xu, R.X., Fesik, S.W. (1992) A 4D HCCH-TOCSY experiment for assigning the side chain 1H and 13C resonances of proteins. *J Biomol NMR* **2**:655-659.

Pope, M. K., Green, B. & Westpheling, J. (1998). The *bldB* gene encodes a small protein required for morphogenesis, antibiotic production, and catabolite control in *Streptomyces coelicolor*. *J Bacteriol* **180**, 1556-1562.

Ryding, N. J., Kelemen, G. H., Whatling, C. A., Flardh, K., Buttner, M. J. & Chater, K. F. (1998). A developmentally regulated gene encoding a repressor-like protein is essential for sporulation in *Streptomyces coelicolor* A3(2). *Mol Microbiol* **29**, 343-357.

Ryding, N. J., Bibb, M. J., Molle, V., Findlay, K. C., Chater, K. F. & Buttner, M. J. (1999). New sporulation loci in *Streptomyces coelicolor* A3(2). *J Bacteriol* **181**, 5419-5425.

Shen, Y., Delaglio, F., Cornilescu, G., Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR

chemical shifts. *J Biomol NMR* **44**:213-223.

Skjaerbaek, N., Nielsen, K.J., Lewis, R.J., Alewood, P., Craik, D.J. (1997) Determination of the solution structures of conantokin-G and conantokin-T by CD and NMR spectroscopy. *J Biol Chem* **272**:2291-2299

Stoll, K.E., Draper, W.E., Kliegman, J.I., Golynskiy, M.V., Brew-Appiah, R.A., Phillips, R.K., Brown, H.K., Breyer, W.A., Jakubovics, N.S., Jenkinson, H.F., Brennan, R.G., Cohen, S.M., Glasfeld, A. (2009) Characterization and structure of the manganese-responsive transcriptional regulator ScaR. *Biochemistry* **48**:10308-10320.

Tan, H., Yang, H., Tian, Y., Wu, W., Whatling, C. A., Chamberlin, L. C., Buttner, M. J., Nodwell, J. & Chater, K. F. (1998). The *Streptomyces coelicolor* sporulation-specific σ^{WhiG} form of RNA polymerase transcribes a gene encoding a ProX-like protein that is dispensable for sporulation. *Gene* **212**, 137-146.

Tillotson, R. D., Wosten, H. A., Richter, M. & Willey, J. M. (1998). A surface active protein involved in aerial hyphae formation in the filamentous fungus *Schizophillum commune* restores the capacity of a bald mutant of the filamentous bacterium *Streptomyces coelicolor* to erect aerial structures. *Mol Microbiol* **30**, 595-602.

Willey, J., Santamaria, R., Guijarro, J., Geistlich, M. & Losick, R. (1991).

Extracellular complementation of a developmental mutation implicates a small sporulation protein in aerial mycelium formation by *S. coelicolor*. *Cell* **65**, 641-650.

Willey, J., Schwedock, J. & Losick, R. (1993). Multiple extracellular signals govern the production of a morphogenetic protein involved in aerial mycelium formation by *Streptomyces coelicolor*. *Genes Dev* **7**, 895-903.

Wishart, D.S., Sykes, B.D. (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR* **4**:171-180.

Wishart, D.S., Sykes, B.D., Richards, F.M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**:1647-1651.

Wüthrich, K. (1986) NMR of Proteins and Nucleic Acids. *John Wiley & Sons, Inc.*

Zheng, N., Fraenkel, E., Pabo, C.O., Pavletich, N.P. (1999) Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. *Genes Dev* **13**:666-674.

국문초록

BldD 단백질은 토양세균인 방선균 내에 존재하는 단백질로 핵산과 결합하는 기능을 가지며, 167개의 아미노산 서열로 구성되어 있다. 방선균 중의 하나인 *Streptomyces coelicolor*는 외부 환경이나 영양 요소의 이용여부에 따라 포자를 형성한다. 이러한 포자 형성은 우선적으로 기균사 형성의 선행이 필수적이다. 이러한 기균사의 형성에 BldD 단백질의 역할이 상당히 중요하다는 많은 보고가 있어 왔고, 실제로 이 BldD 단백질은 분화와 관련된 *bldN* 이나 *whiG* 과 같은 유전자들을 세포 생장 초기에 억제하는 것으로 밝혀졌다. BldD 단백질은 두 개의 독립된 기능체로 이루어져 있으며, 그 중 아미노 말단에 가까이 위치하는 기능체는 BldD 단백질이 핵산과 결합함에 있어 주된 역할을 하는 것이 본 실험실의 연구를 통해 밝혀졌다. 하지만, 카르복시 말단에 가까이 위치하는 기능체는 BldD 단백질이 기능하는데 있어 어떠한 역할을 하고 있는지에 관한 연구 결과는 현재까지 없는 실정이다. 또한 본 실험실의 연구에 의해 BldD 단백질의 이량체 형성에도 아미노 말단 지역의 기능체가 작용하는 것으로 밝혀졌다. 여러 가지 생리학적 또는 생화학적 방법론을 통해 BldD 단백질의 카르복시 말단 지역 기능체에 대한 다양한 연구가 본 실험실에서 시도

되었으나 그 기능을 정확히 동정할 수 없었다. 따라서, 현재까지 시도되었던 생화학적 방법론이 아닌 생물물리학적 접근이 필요한 것으로 생각되었고, 그 중 용액 상태에서의 단백질의 삼차 구조 연구가 가능한 자기공명장치를 이용하여 BldD 단백질의 카르복실 말단 지역 기능체의 삼차 구조를 밝혀내었다. 놀랍게도 그 삼차 구조는 기존에 밝혀졌던 단백질의 구조와는 상이한 새로운 구조인 것으로 밝혀졌으며, 예상과 달리 핵산과 결합할 수 있는 다른 단백질과 높지는 않으나 상당한 정도의 유사성을 보였다. 하지만, 구조적인 유사성과 달리 BldD 단백질의 카르복시 말단 지역 기능체는 일반적인 핵산 결합 단백질과 대조적으로 음전하의 표면을 가지고 있으며, 이 음전하의 분포로 인해 인산에 의해 강한 음전하를 가지는 핵산과 결합할 수 있는 가능성은 매우 적은 것으로 여겨지며, 이것은 젤 이동성 변이 실험을 통해 확인할 수 있었다. 또한 이 단백질은 잘 보존되었고, 외부로 노출되어 있는 소수성 아미노산 집합체를 가지고 있는 것이 구조를 통해 분석되었고, 이러한 소수성 아미노산 집합체의 존재로 볼 때 BldD 단백질의 카르복시 말단 지역 기능체는 다른 단백질과의 상호작용을 통해 BldD 단백질의 전체 기능을 조절하는 것으로 예상된다.

감사의 글

이제는 기억도 흐릿해질 정도의 많은 시간이 걸려 한참이나 부족하지만 한 편의 논문으로 정리하여 박사학위과정을 마무리하게 되었습니다. 과학자로서의 길을 나아감에 있어 남들보다 조금은 늦게 첫 번째 이정표를 지나며, 한 번 뒤돌아보니 많은 아쉬움도 남지만, 무엇보다 결코 혼자서는 이 길을 걸어오지 못했다는 생각이 강하게 듭니다.

먼저, 많이 부족하고 또한 느리디 느린 저를 끝없는 포용과 관심으로 이끌어 주시고, 진정한 과학의 길에 대한 독보적인 철학이 묻어나는 끊임없는 교육으로 제가 앞으로 걷게 될 길의 든든한 반석을 마련해 주신 은사님 강사욱 선생님께 깊은 감사를 올립니다. 실험을 함에 있어 큰 흐름과 그것을 이루는 작은 물줄기 모두를 두루 살펴 주시며 제게 해주셨던 그 모든 말씀들은 제 삶에 있어 큰 밑거름이 될 것이라 확신합니다. 바쁘신 중에도 제 논문을 면밀하게 살펴주시고, 그에 따른 여러 가지 조언들을 아끼지 않고 해주신 석영재 교수님, 허원기 교수님 그리고 최희정 교수님께도 깊은 감사를 드립니다. 그리고, NMR에 대해 이해조차 제대로 못했던 저에게 많은 시간을 할애하여 설명해주시고, NMR 실험과 분석 전반에 걸쳐 도움 주시는 것을 주저하지 않으시며, 논문 작성에 있어서도 지도와 조언을 끊임없이 해주신 원형식 교수님께 깊은 감사를 올립니다. 또한, 자포자기하며 나락으로 빠져버리려던 저에게 끝없는 조언으로 희망을 갖게 해주셨던 임형순 선생님, 정신 누님께 깊이 감사 드립니다.

생물물리학 연구실이라는 같은 공간에서 생활하며 저에게 알게 모르게 많은 도움을 주었던 선배님들과 후배님들께도 감사를 전하고 싶습니다. 선배님들께서 하시는 말씀은 잘 듣지도 않고 하고픈 데로 하고 살았었던 철없던 저를 넓은 마음으로 이해해 주시고, 많은 조언을 해주셨던 모든 선배님들께 깊이 감사 드립니다. 또한, 그다지 좋은 선배의 모습을 보여주지 못하고 변변한 도움조차 주지 못하면서, 부담스럽고 까다로운 선배의 모습만을 보여 주었으나, 그런 선배를 이해해주고 받아주었던 모든 후배님들께도 깊은 감사를 전합니다.

어떠한 말로도 그 은혜를 다 갚을 길이 없겠지만, 이 길을 걸어오기까지 무엇 하나 제대로 된 모습을 보여드리지 못해도, 항상 저를 믿고 묵묵히 지켜봐 주시며, 뒷바라지 하느라 끝없이 희생하신 아버님, 어머님께 깊이 감사를 올립니다. 못난 사위임에도 불구하고 항상 염려해 주시고, 아내보다 저를 더 많이 챙겨주시며, 또 한 번 부모님의 따뜻함을 느낄 수 있게 해주신 장인어른과 장모님께도 깊은 감사를 드립니다.

아빠가 무슨 일을 하는지 잘 모르겠지만, 논문 준비하는 중에 태어나 고생하면서도, 튼튼하고 씩씩하게 자라며 방긋방긋 웃어주는 아들 준우에게 아빠가 많이 고맙고, 많이 사랑한다고 이야기해주고 싶습니다. 마지막으로 지금의 제가 있고, 앞으로의 제가 있기 위한 반석이자 도달점인 사랑하는 아내 지은에게 말로는 다 표현 못할 고마움을 전합니다.

理學博士學位論文

*Streptomyces coelicolor*에서 전사 조절 인자인 BldD의 carboxy 말단 지역 단백질의 기능 및 구조 분석

Functional and structural analyses of the carboxy-terminal domain of the transcription factor BldD in *Streptomyces coelicolor* A3(2)

2014年 2月

서울大學校 大學院

生命科學部

金 正 穆

*Streptomyces coelicolor*에서 전사 조절 인자인 **BldD**의 **carboxy** 말단 지역 단백질의 기능 및 구조 분석

指導敎授　姜 思 旭

이 論文을 理學博士學位論文으로 提出함

2013年　11月

서울大學校

生命科學部

金 正 穆

金正穆의 理學博士學位論文을 認准함

2013年 12月

委 員 長 _____

副委員長 _____

委　　員 _____

委　　員 _____

委　　員 _____

Functional and structural analyses of the carboxy-terminal domain of the transcription factor BldD in *Streptomyces coelicolor* A3(2)

by

Jeong-Mok Kim

Advisor:

Professor Sa-Ouk Kang, Ph. D.

A Thesis Submitted in Partial Fulfillment

of the Requirements for

the Degree of Doctor of Philosophy

February, 2014

School of Biological Sciences

Seoul National University

ABSTRACT

BldD is a DNA-binding protein with 167 amino acids and acts as a repressor for key developmental genes in *Streptomyces coelicolor*. Although extensive researches have emphasized the importance of BldD in developmental processes of *Streptomyces*, distinct regulatory mechanism of BldD has not been well understood yet. The N-terminal domain of BldD (residues 1-79, BldD-NTD) has clear functions that mediate DNA-binding and dimerization, but the function has not been defined for the C-terminal domain of BldD (residues 80-167, BldD-CTD). Therefore, the function of BldD-CTD could more likely be related with the regulatory mechanism of BldD. In this study, backbone and side-chain NMR assignments of the recombinant BldD-CTD protein could be achieved by a series of NMR experiments on a [^{13}C, ^{15}N]-enriched protein sample. The secondary structure prediction by CSI and TALOS+ analysis using the assigned chemical shift data identified that the BldD-CTD adopts a βααβααα fold. From backbone and side-chain assignments of the recombinant BldD-CTD, NOE cross-peaks assignments were also completed for 3D-structure calculation. The determined solution structure of BldD-CTD is very similar to winged-helix domains in spite of different topology. But, DNA-binding of BldD-CTD is not structurally

favorable because of slightly negative-charged surface and additional helical region. As removal of additional helical region did not show any functional difference to native protein, assessed by gel mobility shift assays and *in vivo* complementation experiments, it is the anionic property of the BldD-CTD that appears to be mainly responsible for its inability to bind DNA. Conserved surface analysis of BldD-CTD revealed that highly conserved hydrophobic patch surrounded by charged residues is located opposite to helix-turn-helix region. These structural features suggest that BldD-CTD constitutes a novel fold of winged-helix domain involved in protein-protein interaction and this interaction could be directly related to the regulatory mechanism of BldD.

Key words: *Streptomyces coelicolor*, Differentiation, BldD, DNA-binding protein, NMR, Solution structure, Winged-helix domain, Protein-protein interaction

CONTENTS

LIST OF TABLES

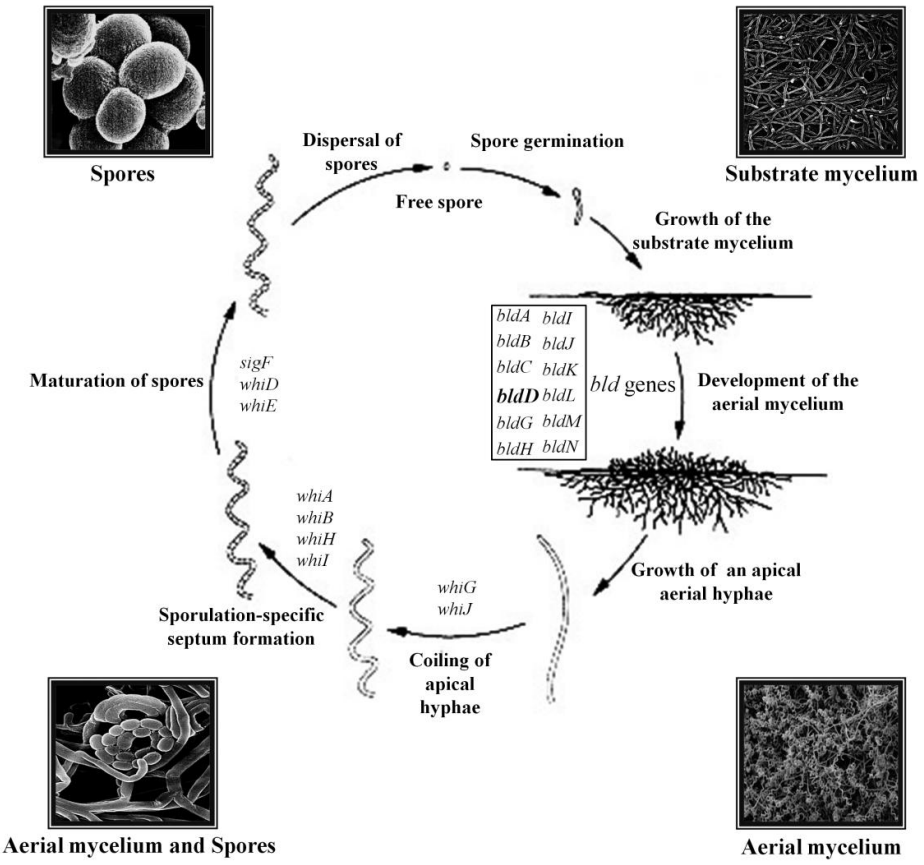
LIST OF FIGURES

I. INTRODUCTION

1. Morphological differentiation of *Streptomyces coelicolor*

Streptomyces coelicolor is spore-forming, filamentous soil bacterium. Spore formation is accomplished through a complex process of morphological differentiation. Under favorable condition, one or two germ tubes emerge from a spore and grow by tip extension and branch formation to give rise to a substrate mycelium. In response to nutrient depletion, aerial hyphae, which grow out into the air constituting a fuzzy layer on the colony surface, are formed and then further differentiate to form chains of unigenomic spores (Kelemen & Buttner, 1998) (Scheme. I). Together with these morphological changes, secondary metabolites such as antibiotics and pigments are also produced (Chater, 1993).

Numerous studies revealed that two classes of genes (*whi* genes and *bld* genes) are essential for the morphological development. The *bld* mutants fail to form aerial mycelium and have a shiny ‘bald’ appearance (Merrick, 1976; Willey *et al.*, 1993). The *whi* mutants are capable of formation of aerial hyphae but lack of the grey spore pigment, of which phenotype yield ‘white’ (Chater, 1972). These mutants also show pleiotropic effects, including defects in cell-



Scheme 1. Life cycle of *Streptomyces coelicolor*. Under favorable condition, one or two germ tubes emerge from a spore and grow by tip extension and branch formation to give rise to a substrate mycelium. In response to nutrient depletion, aerial hyphae, which grow out into the air constituting a fuzzy layer on the colony surface, are formed and then further differentiate to form chains of unigenomic spores The *bld* genes are required for the erection of aerial hyphae and the *whi* genes are required for sporulation (modified from Kieser *et al.*, 2000).

cell signaling and carbon catabolite repression (Kelemen & Buttner, 1998; Merrick, 1976; Willey *et al.*, 1993).

1.1. *whi* mutants

Sporulation-deficient mutants of *S. coelicolor* which showed 'white' phenotype have been mapped genetically, and it has been revealed that all of these mutants have one of the eight separate loci: *whiA*, *whiB*, *whiD*, *whiE*, *whiG*, *whiH*, *whiI*, and *whiJ* (Chater, 1972; Chater & Merrick, 1976; Chater *et al.*, 1989; Davis & Chater, 1992). The *whiA* locus is apparently part of a complex operon and encodes a previously unknown kind of protein with the eukaryotic glutathione peroxidase-related motifs (Ainsa *et al.*, 2000). And *whiB* encodes an unusually small protein (~ 9.9 kDa), the C-terminus of which fulfills some criteria for DNA binding domain that has been suggested for eukaryotic transcription factors (Chater, 1993; Davis & Chater, 1992). Recent growth in the mycobacterial DNA sequence database has led to the discovery of a family of mycobacterial genes encoding WhiB-like proteins. Although the biochemical roles of these homologues are unknown, primary sequence alignment shows that all of them have four conserved cysteinyl residues (CXXC motif), encouraging speculation that they may bind to metals. Perhaps the influence of WhiB on transcription of putative target genes may be responsive to changes in

redox potential (Chater, 1993). WhiD is 112 amino acids long and is a member of a family of proteins that includes WhiB (Molle *et al.*, 2000). The *whiD* mutant formed sporulation septa but failed to go on to produce mature spores. The *whiD* mutant formed spores at wild type abundance but these were unpigmented, were thin-walled, and showed frequent lysis. The *whiE* genes encode proteins that closely resemble the components of type II polyketide synthases, which are involved in the synthesis of a variety of aromatic antibiotics, including tetracenomycin from *Streptomyces glaucescens*, granaticin from *Streptomyces violaceoruber*, oxytetracycline from *Streptomyces rimosus*, and actinorhodin from *S. coelicolor* itself (Kelemen *et al.*, 1998). The sigma factor encoded by *whiG* belongs to flagellar group of sigma factors (Chater *et al.*, 1989; Tan *et al.*, 1998). Since other flagellar group sigma factors such as *B. subtilis* σ^{D} and *E. coli* σ^{F} (*FliA*) are regulated by anti-sigma factor, σ^{whiG} is possibly controlled also by an anti-sigma factor that might be responsive to an event associated with the emergence of aerial hyphae (Losick & Shapiro, 1993). The observed transcription of *whiG* in liquid cultures and early surface cultures devoid of aerial hyphae implies that σ^{whiG} sigma factor is itself present before formation of aerial hyphae (Kelemen *et al.*, 1996). The WhiH protein resembles members of GntR family (FadR, LatR and RdhR) repressors (Ryding *et al.*, 1998). Many members of this family respond to

organic acids that occur in intermediate metabolism. This suggests that WhiH may sense one of these molecules whose concentrations fluctuate significantly in different metabolic states. *In vitro* and *in vivo* evidences strongly indicate that the *whiH* promoter is transcribed by the σ^{whiG}-containing RNA polymerase (Ryding *et al.*, 1998). The onset of *whiH* transcription also coincides with perceptible aerial growth and is much later than that of *whiG* transcription (Ryding *et al.*, 1998). DNA that complements several *whiI* (closely linked to *cysD*) mutants encodes a member of the response regulator family of proteins (Ryding *et al.*, 1999). This gene has putative promoter sequences similar to those of the σ^{whiG}-dependent promoters. The deduced protein of this gene has all the conserved features needed; the adjacent gene appears to encode a membrane-located enzyme of this kind. If this is indeed the *whiI* locus, the membrane-associated kinase domain may permit it to detect signals in or outside of the membrane.

1.2. *bld* genes

At least 10 *bld* loci, *bldA, bldB, bldC, bldD, bldF, bldG, bldH, bldI, bldK, bldL,* have been found in *S. coelicolor* (Champness, 1988; Chater & Merrick, 1976; Merrick, 1976; Nodwell *et al.*, 1996; Willey *et al.*, 1993; Willey *et al.*, 1991). The highly pleiotropic phenotype of these mutations suggests that

these genes are involved with an early stage on the initiation of development. Interestingly, when certain pairs of *bld* mutants are grown on rich media in close proximity, one mutant can trigger the aerial hyphae formation of the other mutant (Molle & Buttner, 2000; Nodwell *et al.*, 1999; Tillotson *et al.*, 1998; Willey *et al.*, 1993; Willey *et al.*, 1991). As extracellular complementation among *bld* mutants is always unidirectional, this can be explained by the hierarchical cascade of intercellular signals for the formation of aerial hyphae. That is, one mutant that is higher in the hierarchy and acts as a donor can give the signals to the other mutant acting as a recipient through unidentified signaling mechanism. Such analyses have suggested the following hierarchy of extracellular complementation groups:

[bldJ] < [bldK – bldL] < [bldA – bldH] < [bldG] < [bldC] < [bldD – bldM]

All blocked mutants to the left are complemented by those to the right, and those in the same group do not interact with each other and display the same pattern of complementation. These mutants are known to be unable to produce a small hydrophobic molecule, SapB that contributes to the erection of the aerial hyphae by reducing the surface tension at the colony surface (Tillotson *et al.*, 1998; Willey *et al.*, 1991). Therefore, in the extracellular complementation of *bld* mutants, one donor mutant renders the recipient mutant to restore the ability to produce the final product, SapB.

Only two *bld* genes, *bldA* and *bldK*, have been characterized at the molecular level. The *bldA* alleles reside in a gene for a leucyl-tRNA that recognized the UUA codon (Lawlor *et al.*, 1987). UUA codon is a rare codon in *Streptomyces*, and it has been suggested that this tRNA is involved in the translation of regulatory genes involved in antibiotic production and morphogenesis (Fernandez-Moreno *et al.*, 1991; Leskiw *et al.*, 1991). It has been shown that the *bldK* locus consists of five adjacent ORFs that specify homologues of the subunits of the oligopeptide permease family of ATP-binding cassette (ABC) membrane-spanning transporters (Nodwell *et al.*, 1996). One of the most intriguing but poorly understood aspects of the *bld* phenotype is that growth on poor carbon sources is sufficient to restore partially the morphological and antibiotic defects of most of these mutants. Both genes for *bldB* and *bldD* were isolated and characterized as small proteins with helix-turn-helix signature of LysR family regulatory proteins (Elliot *et al.*, 1998; Pope *et al.*, 1998), but the actual roles in aerial mycelium formation had remained unknown.

2. General description of *bldD*

Among the *bld* mutants, the *bldD* mutant has mostly severe pleiotropic defects (Merrick, 1976). On minimal medium containing glucose as carbon

source, the *bldD53* mutant has a fragmented surface lacking aerial structures on its colonies and produces none of the four well-characterized antibiotics known to be produced by *S. coelicolor*. The morphological defect of mutant is the case for many of the *bld* mutants, but the loss of antibiotic production is not. Its defect is overcome by growth on minimal medium containing mannitol as the carbon source.

The *bldD* gene product, BldD, consists of 167 amino acids with a calculated molecular weight of 18,167 Da and the identified *bldD* mutant has a point mutation at position 62 from Tyr to Cys (Elliot *et al.*, 1998). BldD binds to its own promoter (Elliot & Leskiw, 1999) and regulatory regions of key developmental genes (Elliot *et al.*, 2001) including *bldN*, which encodes a sigma factor involved in aerial hyphae formation (Bibb *et al.*, 2000), and *whiG*, which encodes a sigma factor related in spore formation (Chater *et al.*, 1989). The stress-responsive genes, *sigH*, is also regulated by BldD, which indicates that BldD is a key regulator connecting stress to developmental processes (Kelemen *et al.*, 2001). The DNaseI footprinting analyses for some of the BldD target genes suggested the imperfect inverted repeat, AGTgA $(n)_m$ TCACc, as the consensus sequence for BldD binding (Elliot *et al.*, 2001). The locations of these binding sites and the differences of transcriptional levels for the putative BldD target genes between wild type and *bldD* mutants revealed that BldD acts

as a repressor during vegetative growth (Elliot & Leskiw, 1999; Elliot *et al.*, 2001). Recently, chromatin immunoprecipitation-microarray analysis (ChIP-chip) has further extended the range of BldD-regulated genes, strengthening the importance of BldD in the life cycle of *Streptomyces coelicolor* (den Hengst *et al*., 2010). But, how BldD regulates various target genes in response to developmental signal is still unknown.

3. Aims of this study

BldD exists predominantly as a homodimer in solution (Elliot *et al*., 2003) and each subunit is composed of two structurally independent domains (Lee *et al*., 2007). The N-terminal domain (residues 1-79, BldD-NTD) is clearly responsible for DNA-binding through its helix-turn-helix motif belonging to XRE (Xenobiotic Response Element) family (Kim *et al*., 2006; Lee *et al*., 2007). Contrarily, the function of the C-terminal domain (residue 80-167, BldD-CTD) has not been understood yet, except that it has a helical structure and acts as a monomer in solution (Lee *et al*., 2007). In the intact BldD dimer, no significant interdomain interaction has been observed between the two distinct domains. Furthermore, isolated BldD-NTD showed no functional defect in dimerization and DNA-binding (Lee *et al*., 2007). Many times of attempts to develop the single crystal from the full-length BldD has ended in failure due to degradation

of BldD-CTD. Thus, the structural information of BldD-CTD acquired by NMR in solution is essential to understand the precise regulatory action of BldD directly linked to the developmental processes in *Streptomyces coelicolor*. In this work, almost complete NMR assignments for BldD-CTD were accomplished and three dimensional BldD-CTD structure with high quality was determined. The BldD-CTD structure adopts a novel fold similar to winged-helix domain, but was not compatible with DNA-binding. Therefore, BldD-CTD structure would provide new insight for regulatory mechanism of BldD, and extensively to role of winged-helix domain.

II. Materials and Methods

1. Bacterial strains, media and culture conditions

S. coelicolor A3(2) strain was grown and maintained as described by Hopwood *et al*. (1985). For liquid culture, pre-germinated spores (about 10^8~10^9 spores/100ml broth) were grown in YEME liquid medium (1% glucose, 0.5% Bacto-peptone, 0.3% malt extract (Difco), 0.3% yeast extract (Difco)) containing 34% sucrose at 30°C with vigorous shaking. Pre-germinated spores (about 10^7) or patches of mycelia were inoculated on R2YE (10.3% sucrose, 1% glucose, 1.01% $MgCl_2$, 0.024% K_2SO_4, 0.001% casamino acid (Difco), 0.5% yeast extract, 1.43% (~20 mM) TES (N-tris[hydroxymethyl]methyl-2-aminoenthanesulfonic acid, pH 7.0), 20 mM $CaCl_2$, 0.005% K_2HPO_4, and 0.3% proline) for analysis of phenotype or NA (0.8% nutrient broth (Difco)) agar media for selection of mutant strains or MS (2% D-mannitol, 2% soybean flour and 2% agar) for spore preparation and grown at 30°C. For preparation of mycelia on solid media, 10^7 pre-germinated spores were inoculated on R2YE over laid with cellophane discs. The morphological developmental process was confirmed routinely by microscopic observations. All the recombinant DNAs were introduced into competent *E. coli* DH5α (Sambrook *et al*., 1989). For

conjugation of DNA into *S. coelicolor*, methylation-deficient *E. coli* ET12567 containing pUZ8002 which is a compatible *oriT*-containing plasmid. For overexpression of recombinant proteins using T7 polymerase-based system, *E. coli* BL21(DE3)pLysS (Novagen) was used. *E. coli* cells were grown at 37°C in LB (1% tryptone (Difco), 0.5% yeast extract, and 1% NaCl) or M9 (1.79% $Na_2HPO_4 \cdot 12H_2O$, 0.3% KH_2PO_4, 0.1% NH_4Cl, 0.1 mM $CaCl_2$, 0.2 M $MgSO_4$, 1 μg·mL^{-1} Thiamine, 0.3% Glucose) supplemented with appropriate antibiotics.

2. Construction of plasmids for overexpression of BldD, BldD-CTD, and their truncated forms

Oligonucleotide primers were designed from the sequence of *bldD* coding region with restriction sites including NdeI and BamHI. DNA were amplified by PCR (Polymerase Chain Reaction) using *S. coelicolor* A3(2) chromosomal DNA as a template. PCR reactions were performed in a Perkin-Elmer Cetus thermal cycler for 30 cycles with the following optimized conditions: denaturation at 98°C for 20 seconds, annealing at 68°C for 20 seconds, extension at 72°C for 1 min. Amplified DNA were separated by 1% agarose gel and visualized on UV-illuminator. The DNA band matching to expected size was sliced, extracted and cleaned by gene cleaning kit (MoBio).

Purified DNAs were introduced into pGEM-Teasy vector (Promega) and were sequenced for the verification of amplified sequences. The pGEM-Teasy vector containing verified gene coding region corresponding to BldD or BldD-CTD was excised by restriction enzymes, NdeI and BamHI. The excised DNA was purified and cleaned with same procedures for PCR products. The pET-15b vector (Novagen) was also excised by same restriction enzymes, purified, and cleaned with same procedures. The resultant excised pET-15b vector and DNA corresponding to coding region for BldD or BldD-CTD were ligated by T4 ligase (Fermentas) following provided protocol. The plasmids for the truncated forms of BldD or BldD-CTD were also constructed following same procedures for construction of plasmid containing coding region of BldD or BldD-CTD.

3. *In vivo* complementation experiments

To construct a plasmid for complementation experiment, the DNAs containing promoter region of *bldD* and ORF for BldD and its fragments were amplified with PCR and cloned into pGEM-Teasy vector. DNA fragments from digestion of EcoRI were ligated to into pSET162, which has an insertion of a thiostrepton resistance marker at SphI site of pSET152 (Bierman *et al*., 1992). The constructed pSET162 vectors for *in vivo* complementation were verified by sequencing the purified vectors. The pSET162 derivatives were introduced into

methylation-deficient ET12567 containing the pUZ8002 vector for conjugation. The *bldD* deletion mutant was conjugated with the transformed cells derived from ET12567/PUZ8002. The integration of recombinant construct to *bldD* deletion mutant was verified by resistance for apramycin and thiostrepton on NA solid medium and PCR analysis of genomic DNA from the pSET162-integrated candidates. For phenotypic analysis, wild type and mutant strains were grown on R2YE solid medium at 30°C and monitored after 5 days.

4. Overexpression and purification of BldD, BldD-CTD, and their truncated forms

A freshly prepared *E. coli* BL21(DE3)pLysS was transformed with pET-15b or pET-3a vector containing sequence corresponding to BldD or BldD-CTD or its fragments. The transformed cells were grown in LB medium containing 50 μg/ml ampicillin and 30 μg/ml chloramphenicol at 37°C. When the cell density reached an OD_{600} of about 0.5, protein expression was induced with 1 mM IPTG and cells were incubated at 22°C for 15 h. The cells were harvested by centrifugation, resuspended in 100 ml binding buffer containing 1 mM PMSF and disrupted by sonication. After centrifugation for removing debris of cells and insoluble fraction, the supernatant was applied to a nickel affinity column. The column was washed with 5 bed volume of binding buffer

(5 mM imidazole, 100 mM NaCl, and 10 mM Tris-HCl, at pH 7.9) and sequentially with 10 bed volume of washing buffer (25 mM imidazole, 100 mM NaCl, and 10 mM Tris-HCl, at pH 7.9). The proteins were eluted with 3 bed volume of elution buffer (500 mM imidazole, 100 mM NaCl, and 10 mM Tris-HCl, at pH 7.9). The eluent from Ni-affinity column was concentrated by ultrafiltration using YM-1 membrane (Millipore). The concentrated protein was applied to PD-10 column (GE healthcare) for exchanging buffer to thrombin cleavage buffer. The eluent in thrombin cleavage buffer (20 mM Tris-HCl buffer containing 150 mM NaCl and 2.5 mM $CaCl_2$, at pH 8.4) was mixed with 2 unit of thrombin (Novagen) and incubated at 22°C for 18 h to cleave the N-terminally tagged histidines, followed by the removal of thrombin and other impurities via the sequential application of nickel-affinity. For further purification of proteins and buffer-exchange, eluent was applied to gel-filtration chromatography (HiLoad 16/60 Superdex^{TM} 75, GE healthcare) and concentrated in 20 mM Tris-HCl buffer (pH 7.8) containing 150 mM NaCl. Purified proteins were qualified with tris-tricine polyacrylamide gel electrophoresis and quantified with Protein assay kits (BioRad) based on the method by Lowry.

5. Immunoblot analysis of *S. coelicolor*

The collected mycelia were resuspended in 20 mM Tris-HCl buffer (pH 7.8) containing 150 mM NaCl, 2 mM EDTA, 1 mM PMSF and 1% (v/v) Triton X-100. The resuspended mycelia were disrupted by sonication and clarified by centrifugation at 4°C. The concentration of the total protein in the crude cell extract or purified proteins was accurately quantified using Protein assay kits (BioRad) based on the method by Lowry. The 40 μg of crude extract was applied on 15.4% Tris-Tricine polyacrylamide gel electrophoresis with constant current (30 mA). The polyacrylamide gel was transferred to PVDF (Millipore) membrane using TransBlot system (BioRad) at 150 mA for 90 min. The transferred membrane was washed and blocked with three times of 10 ml Tris-buffered saline (TBS; 10 mM Tris-HCl, at pH 7.5, and 150 mM NaCl) containing 0.05% Tween20 (TBST) with 1% skim milk (Difco). Membranes were incubated with anti-BldD-NTD antiserum (from rabbit) or anti-FLAG-M2 monoclonal antibody (sigma) in TBST containing 1% skim milk for 12 hours (ratio 1:2000 for anti-BldD-NTD and 1:10000 for anti-FLAG-M2). Non-specific bound or unbound antibodies were removed by repeated washing (at least three times) with TBST. The membrane was incubated for 60 min with secondary antibody solution (anti-rabbit or anti-mouse IgG antibody-alkaline phosphatase conjugate, diluted to 1:10000 in TBST). The membrane was then washed for 15 min with three changes of TBST and rinsed with alkaline

phosphatase buffer (100 mM Tris-HCl, at pH 9.5, 100 mM NaCl, and 10 mM $MgCl_2$). The membrane was visualized through incubation in 10 ml of alkaline phosphatase buffer containing 8 mg of BCIP and 16 mg of NBT at RT.

6. Gel mobility shift assays

To prepare the probes for gel shift assay, complementary oligonucleotides were synthesized as follows:

p*whiG*I-F 5′(6-FAM) AAGGTGTTCGAGTGATCACCCAGAGCGA 3′

p*whiG*I-R 5′ ATCGCTCTGGGTGATCACTCGAACACCT 3′

p*whiG*II-F 5′(6-FAM) AAGTCCAGTCACGCTACGCTCACGATGA 3′

p*whiG*II-R 5′ ATCATCGTGAGCGTAGCGTGACTGGACT 3′

p*bdtA*-F 5′(6-FAM) AGCACGCAGCGACGAAGAGTCACCGGAA 3′

p*bdtA*-R 5′ ATTCCGGTGACTCTTCGTCGCTGCGTGC 3′

p*bldN*I-F 5′(6-FAM) ACAGTGCCTGCACGAAGCGTTATTCTCCT 3′

p*bldN*I-R 5′ AAGGAGAATAACGCTTCGTGCAGGCACTG 3′

p*bldN*II-F 5′(6-FAM) ACGGGTGAATGGTTCCGTACTGCACGTG 3′

p*bldN*II-R 5′ ACACGTGCAGTACGGAACCATTCACCCG 3′

p*bldD*-F 5′(6-FAM) AAGCAGAGTAACGCTGCGTAACCTCACA 3′

p*bldD*-R 5′(6-FAM) ATGTGAGGTTACGCAGCGTTACTCTGCT 3′

Each oligonucleotide of forward direction (annotated as F) was labeled with 6-FAM to its 5′ end. These oligonucleotides were annealed in an annealing buffer containing 20 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 50 mM NaCl. The annealing mixture was heated to 100°C for 5 min and then slowly cooled over 80 min to 30°C. These annealed oligonucleotides were precipitated with 1/10 volume of $LiCl_2$ and 2 volume of absolute ethanol, followed by centrifugation and washing pellets with 70% ethanol. The annealed probes were qualified through 8% native PAGE and quantified with absorbance at 260 nm.

Gel mobility shift assays were performed using 6-FAM-labeled probes and purified proteins. The BldD, BldD-CTD, and their truncated forms were incubated with 6-FAM-labeled probe at 30°C for 20 min in binding buffer consisting of 10 mM Tris-HCl (pH 7.8), 150 mM NaCl, 2 mM dithiothreitol, 1 μg of poly (dI-dC), and 10% glycerol. The DNA-protein complexes were run out on an 8% native polyacrylamide gel for 1 hrs at 20 mA. The gel was visualized with 60s exposure time in SYBR mode using LAS3000 (Fujifilm).

7. NMR spectroscopy

7.1 Stable isotope labeling

For uniformly labeling of BldD-CTD (residues 80–167) with stable

isotope, *E. coli* BL21(DE3)pLysS cells containing pET-15b vector for expression of BldD-CTD were grown at 37 °C in M9 minimal media supplemented with $^{15}NH_4Cl$ and ^{13}C-glucose as the sole nitrogen and carbon sources, respectively. The same procedures for expression and purification of proteins were performed as mentioned above in section 4. Finally, 2 mM of uniformly [^{13}C,^{15}N]-enriched BldD-CTD, dissolved in a 40 mM sodium phosphate buffer at pH 6.8 and containing 93% H_2O/7% D_2O, 150 mM NaCl, and 0.05% (w/v) NaN_3, was prepared for NMR.

7.2 NMR measurement and chemical shift assignment

NMR experiments were performed at 295 K on a Bruker Biospin Avance 500 NMR spectrometer equipped with a cryoprobe at KBSI (Korea Basic Science Institute, Chungbuk, Korea). 2D-[^{1}H,^{15}N]HSQC, 2D-[^{1}H,^{13}C]HSQC, 3D-HNCACB, 3D-HN(CO)CACB, 3D-HNCO, 3D-^{15}N-TOCSY-HSQC, 3D-HCCH-COSY, and 3D-HCCH-TOCSY spectra were measured for backbone and side-chain resonance assignments. For referencing the chemical shifts, resonance frequency of DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) was measured. Chemical shifts of ^{1}H atoms were referenced directly, and those of ^{15}N and ^{13}C atoms were referenced indirectly, using the chemical shift ratio values suggested in the BMRB (Biological Magnetic

Resonance Bank: http://www.bmrb.wisc.edu).

All NMR spectra were processed using the NMRPipe (Delagio *et al*., 1995) software package and analyzed with NMRView (Johnson, 2004) program. In processing the raw FIDs (Free Induction Decay), to minimize water signal, the solvent suppression filter was applied to ^{1}H dimension of all data sets (Marion *et al*., 1989). Two-fold zero-filling was also applied to each dimension for fine resolution of data. 3D-[^{15}N]-edited NOESY and 3D-[^{13}C]-edited NOESY spectra were assigned by manual and automated assignments using CYANA 2.1 (Herrmann *et al*., 2002) program.

7.3 Secondary structure prediction and 3D-strucrue calculation

Secondary structure elements was predicted by CSI (Chemical Shift Index) (Wishart et al., 1992; Wishart et al., 1994) and TALOS+ analysis (Shen et al., 2009) with the assigned $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$ and $^{1}H^{\alpha}$ chemical shifts. Dihedral angle restraints derived from the prediction of TALOS+ and distance restraints from manual assignments were used for initial calculation of structure. Initial structures of the BldD-CTD from manual assignment were calculated only by torsion angle dynamics algorithm DYANA (Herrmann *et al*., 2002) included in CYANA program. For automated assignment of NOESY signal, CANDID module was used with initial input of manual assignments and dihedral angle

restraints. The NOESY cross-peaks and its distance restraints from automated assignment were validated by thoroughly examination of visualized each NOESY spectrum. From repetitive validation of automated assignment, almost all of the noise peaks and wrong-assigned peaks are removed or corrected. The 20 structures with the lowest target function values from the final cycle of CYANA calculation were further refined with explicit water (Linge *et al*., 2003) using CNS (Brünger *et al*., 1998) with RECOORD script (Nederveen *et al*., 2005).

7.4 Validation and deposition of calculated structures

The final 20 structures out of total 100 structures calculated were validated using the PSVS 1.5 software package (Bhattacharya *et al*., 2007). Atomic coordinates for the 20 structures with lowest energies and all assigned chemical shift values were deposited in the Protein Data Bank (PDB ID 2MC4) and Biological Magnetic Resonance Bank (BMRB accession number 19427), respectively.

7.5 Heteronuclear NOE measurement

The steady-state [^{1}H,^{15}N]-heteronuclear NOE values were acquired from the [^{1}H,^{15}N]-HSQC spectra recorded with (NOE experiment) and without

(NONOE experiment) using a proton presaturation period of 3 sec (Farrow *et al*., 1994). The NOE values of individual residues were determined from the ratio of the peak intensities with and without saturation.

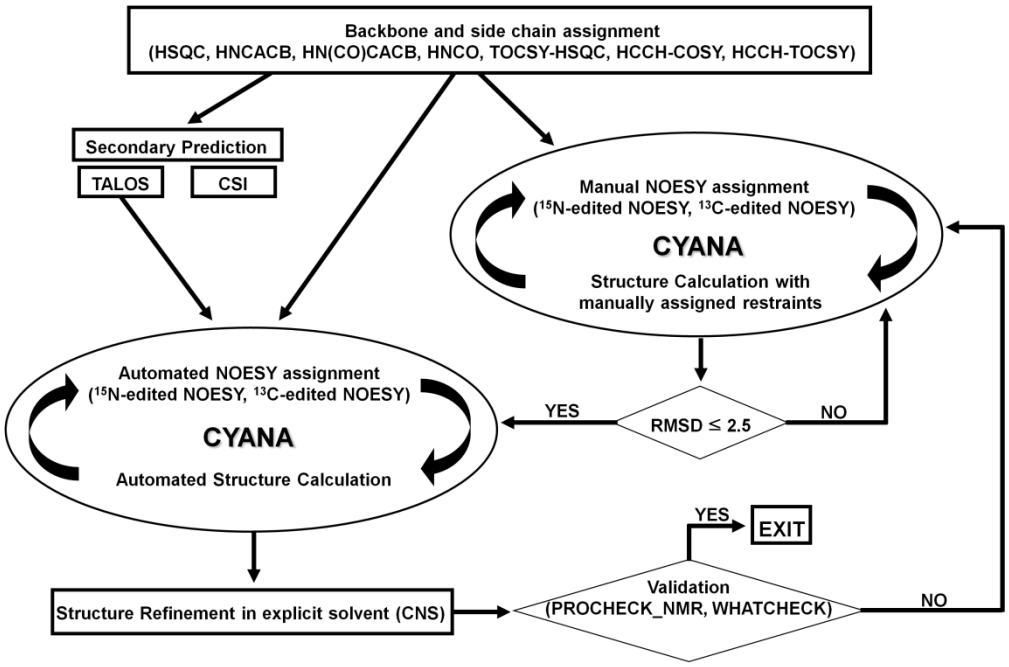


Fig. 1. Overall scheme of solution structure calculation. Based on backbone and side-chain assignment, NOE cross-peaks are assigned manually or automatically. If structure generated from manual assignment has a RMSD value below 2.5 Å, distance restraints of manual assignment can be included in automated structure calculation process of CYANA. Dihedral angle restraints from TALOS are also used in CYANA as NOE-independent restraints. To bring protein structure close to physical reality, structure generated from CYANA goes through refinement process. Final structure is validated to evaluate structure quality.

III. Data and Results

The solution 3D-structure calculation from NOESY cross-peaks has several processes summarized in Fig. 1 (Güntert, 2004). The assignment of chemical shifts using spectra from NMR measurements of proteins is the first step for analysis of NOESY cross-peaks.

1. Verification of peak clusters for backbone assignments

2D-[^{1}H/^{15}N]HSQC, 3D-^{15}N-TOCSY-HSQC and a series of triple resonance spectra including 3D-HNCACB, 3D-HN(CO)CACB, 3D-HNCO were measured on uniformly [^{13}C/^{15}N]-labeled BldD-CTD. As 2D-[^{1}H/^{15}N]HSQC provides the correlation between $^{1}H^{N}$ and ^{15}N, one peak in the spectrum corresponds to one residue from proteins. In a peak from triple resonance spectra, correlation between ^{1}H and ^{15}N also matched to one residue from proteins (Kay *et al*., 2011). Therefore, based on peaks of 2D-[^{1}H/^{15}N]HSQC spectrum, peaks on the each spectrum can be clustered as shown in Fig. 2.

In the HN(CO)CACB scheme, only the inter-residue correlations of the form {$^{13}C^{\alpha}$(i-1), $^{13}C^{\beta}$(i-1), ^{15}N(i), $^{1}H^{N}$(i)} are observed, while both intra- and

inter-residue correlations of the form {$^{13}C^{\alpha}$(i-1), $^{13}C^{\beta}$(i-1), $^{13}C^{\alpha}$(i), $^{13}C^{\beta}$(i), ^{15}N(i), $^{1}H^{N}$(i)} are observed in the HNCACB scheme (Kay *et al*., 2011). In case of the HNCO, only $^{13}C'$(i-1), ^{15}N(i), $^{1}H^{N}$(i) correlations are observed. 3D-^{15}N-TOCSY-HSQC shows correlation between amide proton and all side-chain protons from one residue of proteins (Marion *et al*., 1989). From the 3D-^{15}N-TOCSY-HSQC spectrum, $^{1}H^{N}$-^{15}N-$^{1}H^{\alpha}$ correlations are observed. As a result of peak clustering, one peak cluster has a information of chemical shifts set {$^{1}H^{N}$(i), ^{15}N(i), $^{13}C^{\alpha}$(i), $^{13}C^{\beta}$(i), $^{13}C^{\alpha}$(i-1), $^{13}C^{\beta}$(i-1), ^{13}CO(i-1), $^{1}H^{\alpha}$(i)}.

2. Sequential assignments by linking peak clusters

Sequential backbone assignments of BldD-CTD could be accomplished by sequential linking of the peak clusters identified. As shown in Fig. 3, strip plots of the 3D-HNCACB spectra presented for residues from T109 to Q118 is sequentially linked from the correlation $^{1}H^{N}$(i)-^{15}N(i)-$^{13}C^{\alpha}$(i)-$^{13}C^{\beta}$(i)-$^{13}C^{\alpha}$(i-1)-$^{13}C^{\beta}$(i-1). As a result from sequential backbone linking procedure, the assignment of the 2D-[^{1}H/^{15}N]HSQC is shown in Fig. 4 and assigned chemical shifts of $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, ^{13}CO, $^{1}H^{\alpha}$ are summarized in Table 1. In the 2D-[^{1}H/^{15}N]HSQC spectrum (Fig. 4), the seven pairs of NH_2 signals from six glutamines and one asparagine could be clearly distinguished from the

backbone amide resonances (indicated by lines in Fig. 4). However, a few signals without labeling in Fig. 4, which might originate from the non-assigned backbone amides or from the arginine guanidine groups, could not be assigned due to the absence of corresponding signals in the 3D NMR spectra. Finally, extents of the present sequence-specific assignments of native sequence correspond to 97.6% for $^{1}H^{N}$, 90.9% for ^{15}N, 96.6% for $^{13}C^{\alpha}$, 95.3% for $^{13}C^{\beta}$, and 90.8% for $^{13}C'$ atoms. Then, total 85 of $^{1}H^{\alpha}$ atoms were unambiguously assigned based on the $^{1}H^{N}$-^{15}N-$^{1}H^{\alpha}$ correlations in 3D-^{15}N-TOCSY-HSQC spectra.

3. CSI and TALOS+ prediction of BldD-CTD

The CSI (Wishart et al., 1992; Wishart et al., 1994) method and TALOS (Cornilescu *et al*., 1999) were developed to predict a secondary structure based on chemical shifts of protein backbone atoms. Because of dependency on backbone atoms, CSI method and TALOS can also provide a NOE-independent structure prediction. Since the CSI method is a statistical technique, and the TALOS-derived ϕ/φ-values are empirical in nature, these approaches to structural information of proteins complement each other.

In the CSI method, the mark ‘1’ represents the α-helix tendency of the

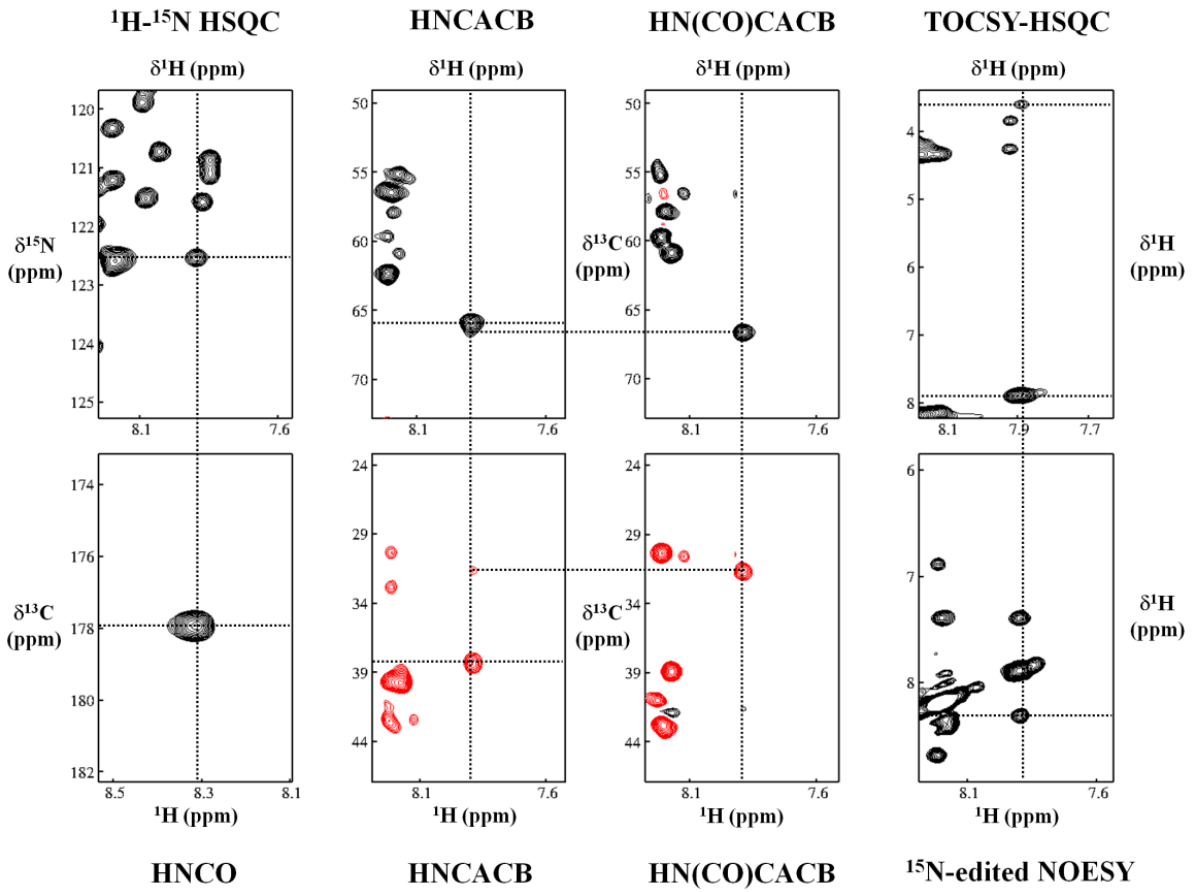

^{1}H-^{15}N HSQC
HNCACB
HN(CO)CACB
TOCSY-HSQC
δ^{1}H (ppm)
δ^{15}N (ppm)
δ^{13}C (ppm)
^{1}H (ppm)
HNCO
HNCACB
HN(CO)CACB
^{15}N-edited NOESY

Fig. 2. Verification of a peak cluster of Ile135 for backbone assignment. The 8 peaks are combined into a cluster. The X-axis of the individual spectrum represents the ^{1}H chemical shifts. The Y-axis of the 2D-[$^{1}H/^{15}N$]HSQC spectrum represent $^{15}N^{H}$ chemical shifts, and those of the 3D-^{15}N-TOCSY-HSQC and 3D-[^{15}N]-edited NOESY spectra represents ^{1}H chemical shifts. The Y-axis of the other spectra represents the ^{13}C chemical shifts of the same $^{15}N^{H}$-edited slide. Each peak at the center of the crosshair in each spectrum appears with the same ^{1}H and $^{15}N^{H}$ chemical shifts.

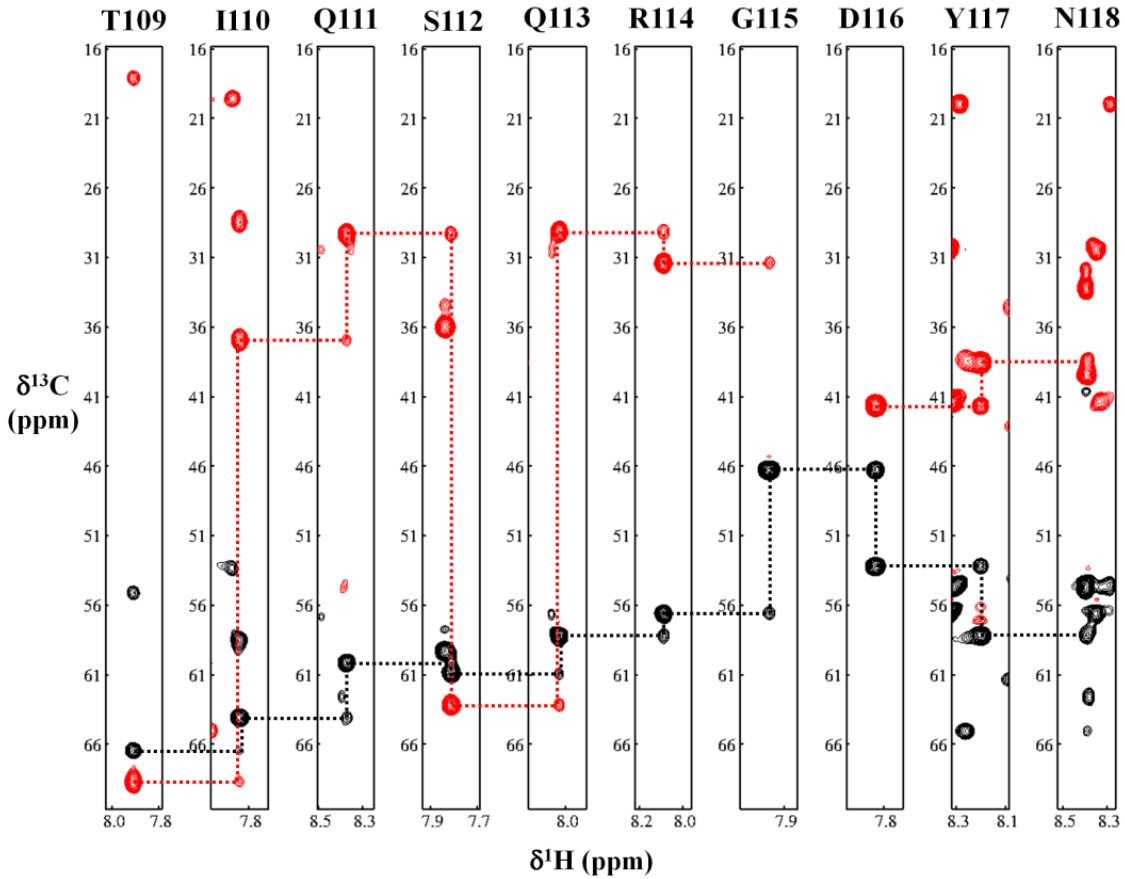

T109
I110
Q111
S112
Q113
R114
G115
D116
Y117
N118
δ13C (ppm)
δ1H (ppm)

Fig. 3. Sequential linking of peak clusters from T109 to Q118. The 10 strip plots of 3D-HNCACB are visualized for sequential linking. As a correlation of $^{1}H^{N}$(i)-^{15}N(i)-$^{13}C^{\alpha}$(i)-$^{13}C^{\beta}$(i)-$^{13}C^{\alpha}$(i-1)-$^{13}C^{\beta}$(i-1), $13C^{\alpha}$ (black dotted line) and $13C^{\beta}$ (red dotted line) are sequentially linked through the plots. The broken linking trace line shows the glycine residue (G115) which has not C^{β}.

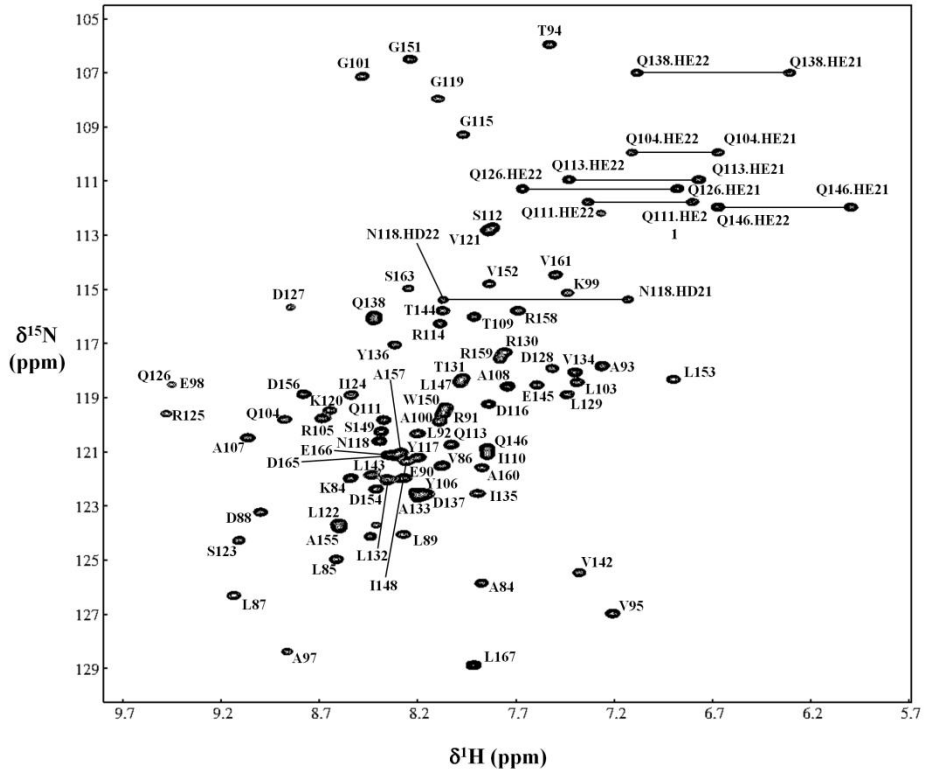


Fig. 4. 2D-[^{1}H/^{15}N]HSQC spectrum of BldD-CTD. Each correlation in the spectrum is labeled with the corresponding residue finally assigned. Side-chain signals from glutamine and asparagine residues are also labeled from assignment of 3D-[^{15}N]-edited NOESY spectrum and indicated by horizontal lines.

Table 1. Backbone assignments of BldD-CTD, measured at 295 K and pH 6.8 (NA, not available; ND, not detected).

Residue	$^{1}H^{N}$	$^{15}N^{H}$	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C'$	$^{1}H^{\alpha}$
E80	ND	ND	ND	ND	ND	ND
P81	NA	NA	ND	ND	ND	ND
P82	NA	NA	ND	ND	ND	ND
P83	NA	NA	62.78	32.35	176.3	4.401
K84	8.542	122.0	56.17	33.27	175.7	4.374
L85	8.615	125.0	54.05	43.17	175.0	4.523
V86	8.078	121.5	61.21	34.60	175.0	4.547
L87	9.136	126.3	56.55	44.23	175.8	4.940
D88	9.000	123.2	53.59	40.82	176.7	4.848
L89	8.272	124.0	57.01	40.85	180.2	4.066
E90	8.270	122.0	59.63	29.56	179.7	4.174
R91	8.075	119.6	57.31	29.56	179.9	4.172
L92	8.200	120.3	57.57	41.80	177.2	3.947
A93	7.262	117.8	54.32	18.71	178.5	4.125
T94	7.529	106.0	61.64	69.96	174.6	4.378
V95	7.206	127.0	60.54	32.39	ND	3.985
P96	NA	NA	63.85	32.71	178.1	4.268
A97	8.865	128.4	55.32	18.91	180.0	3.944
E98	9.453	118.5	59.64	28.84	177.6	4.114
K99	7.437	115.1	56.99	33.95	177.0	4.652
A100	8.096	119.9	53.18	20.23	177.5	4.144
G101	8.482	107.1	49.01	NA	ND	4.138
P102	NA	NA	65.95	31.37	179.3	4.174
L103	7.390	118.4	57.94	42.35	177.9	4.356
Q104	8.879	119.8	59.77	27.55	179.9	3.894
R105	8.688	119.8	59.64	30.42	179.0	4.141
Y106	8.214	122.6	62.28	39.71	177.8	4.171
A107	9.065	120.5	55.19	17.93	178.7	3.762
A108	7.741	118.6	55.08	18.09	180.9	4.287
T109	7.912	116.0	66.45	68.87	176.7	3.995

Table 1. (Continued)

Residue	$^{1}H^{N}$	$^{15}N^{H}$	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C'$	$^{1}H^{\alpha}$
I110	7.844	121.1	64.15	36.94	178.6	3.815
Q111	8.373	119.8	59.86	29.37	178.3	3.839
S112	7.813	112.7	60.92	63.24	176.8	4.333
Q113	8.029	120.7	58.11	29.19	177.9	4.194
R114	8.088	116.3	56.42	31.42	177.0	4.398
G115	7.968	109.3	46.27	NA	173.7	3.880
D116	7.838	119.2	53.17	41.64	175.4	4.715
Y117	8.199	121.2	58.13	38.59	176.9	4.527
N118	8.391	120.6	54.38	39.51	175.9	4.603
G119	8.097	108.0	45.93	NA	173.9	3.929
K120	8.647	119.5	57.71	34.35	176.2	4.489
V121	7.838	112.8	59.26	35.95	174.9	5.330
L122	8.604	123.7	53.88	47.50	175.3	4.789
S123	9.110	124.3	58.85	63.51	174.3	5.052
I124	8.540	118.9	59.60	41.60	174.2	4.780
R125	9.481	119.6	54.51	33.32	178.5	4.865
Q126	9.345	122.8	60.53	28.28	179.0	3.818
D127	8.848	115.7	57.36	40.48	178.8	4.360
D128	7.515	117.9	58.08	41.64	177.9	4.612
L129	7.441	118.9	58.13	41.40	178.3	3.912
R130	7.758	117.3	60.29	29.82	178.9	3.910
T131	7.975	118.3	67.00	68.76	176.6	3.924
L132	8.354	122.0	57.82	43.09	177.8	3.809
A133	8.197	122.6	56.51	17.16	179.2	3.892
V134	7.401	118.1	66.59	31.69	180.4	3.861
I135	7.894	122.5	65.84	38.39	177.9	3.604
Y136	8.317	117.1	60.76	39.03	174.7	4.131
D137	8.173	122.6	55.11	39.81	174.9	4.350
Q138	8.422	116.0	53.51	34.38	174.9	4.719
S139	8.422	116.1	56.44	62.81	ND	4.783

Table 1. (Continued)

Residue	$^{1}H^{N}$	$^{15}N^{H}$	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C'$	$^{1}H^{\alpha}$
P140	NA	NA	66.50	31.78	179.2	4.036
S141	ND	ND	62.37	62.29	176.1	4.838
V142	7.378	125.5	66.43	32.02	178.8	3.561
L143	8.438	121.9	57.02	41.26	178.8	3.899
T144	8.074	115.8	69.06	67.99	175.6	3.321
E145	7.593	118.5	59.23	28.92	179.3	3.841
Q146	7.845	120.9	58.39	28.37	177.4	3.100
L147	7.982	118.4	58.41	40.97	180.2	3.473
I148	8.258	121.4	65.05	38.35	180.9	4.160
S149	8.386	120.3	62.54	62.84	177.0	4.273
W150	8.061	119.4	55.76	31.26	176.8	4.926
G151	8.239	106.5	45.40	NA	174.7	4.167
V152	7.835	114.8	63.11	32.85	174.0	4.197
L153	6.897	118.3	52.63	47.16	175.0	4.911
D154	8.412	122.4	54.38	43.23	176.4	4.603
A155	8.600	123.8	54.73	18.75	179.1	4.054
D156	8.784	118.9	56.28	40.91	174.9	4.513
A157	8.287	121.0	54.39	19.92	178.8	4.001
R158	7.688	115.8	59.23	30.11	178.3	3.924
R159	7.781	117.6	57.69	30.24	177.2	4.206
A160	7.872	121.6	53.26	19.56	177.8	4.279
V161	7.500	114.5	62.04	32.78	175.6	4.141
A162	7.874	125.8	52.77	19.51	177.7	4.321
S163	8.248	115.0	58.39	63.92	ND	4.396
H164	7.971	125.4	56.32	30.31	174.9	4.666
D165	8.322	121.1	54.58	41.26	175.9	4.603
E166	8.344	121.1	56.37	30.37	175.5	4.317
L167	7.916	128.9	57.15	43.45	ND	4.165

atom of the residue, while the mark '-1' represents β-strand tendency. The mark '0' shows that residue does not have a strand or a helical tendency, of which the chemical shift is within the reference value range.

The TALOS is a new database system for the prediction of ϕ and φ backbone torsion angles using a combination of chemical shift assignments for a given protein sequence (Cornilescu *et al*., 1999). The TALOS output for ϕ and φ angles of the center residue in each string consists of the average of corresponding angles in the 10 strings in the database with the highest degree of similarity. The TALOS+ is enhanced version of TALOS which further enhances the prediction rate to 88.5%, without increasing the error rate from addition of a two-layer neural network filter to the database fragment selection process (Shen et al., 2009).

All of the assigned chemical shift data sets were applied to CSI and TALOS+ analysis to predict the secondary structure elements of BldD-CTD (Fig. 5). The determined secondary structure by combining the CSI and TALOS+ results indicated that BldD-CTD is predominantly α-helical, similar to the previous far-UV CD results (Lee et al., 2007). However, the present results additionally revealed the presence of two short β-strands, therby forming a βααβααα topology. The two β-strands are expected to form a β-sheet since

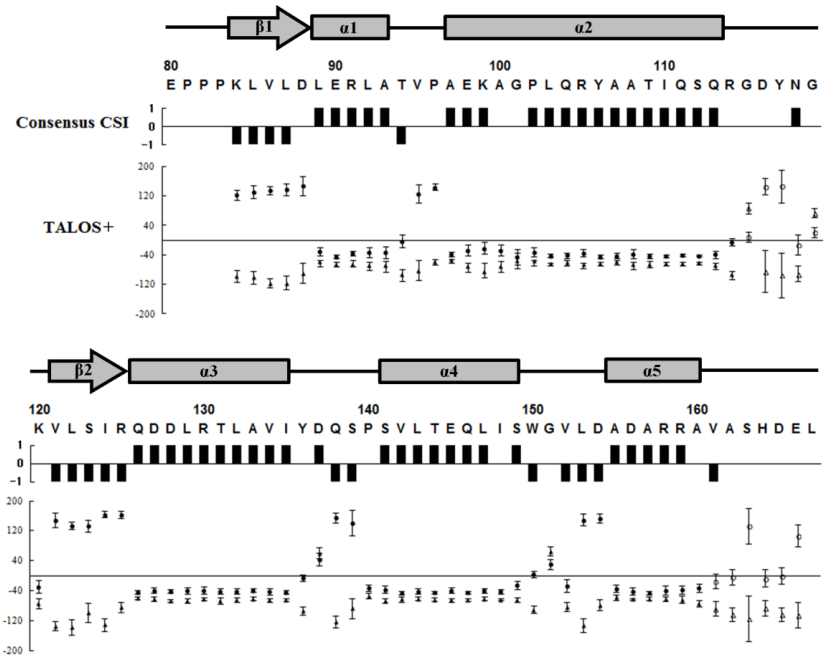


Fig. 5. Secondary structure prediction from CSI methods and TALOS+. In the CSI results, the mark "1" represents the α-helical tendency of the residue (downfield shifts of $^{13}C^{\alpha}$ or $^{13}C'$ resonances, and upfield shifts of $^{13}C^{\beta}$ or $^{1}H^{\alpha}$ resonances from their reference value ranges), while "-1" represents the opposite pattern (β-strand tendency). The chemical shift within the reference value range was marked as a "0". The length of error bars with the TALOS+-predicted backbone dihedral angles, ϕ (triangles) and φ (rectangles), indicate the standard deviation from the average. Predicted secondary structure elements are indicated by boxes for α-helices and arrows for β-strands, along the amino acid sequence.

their lengths are almost the same.

4. Verification of peak clusters for side-chain assignments

Based on backbone assignments, side-chain assignment was performed through analyses of 3D-HCCH-COSY (Ikura *et al*., 1991) and 3D-HCCH-TOCSY (Olejniczak *et al*., 1992). In 3D-HCCH-COSY and 3D-HCCH-TOCSY, magnetisation is transferred from the side-chain hydrogen nuclei to their attached ^{13}C nuclei. While only the hydrogen resonances of the own and within three covalent bond in 3D-HCCH-COSY, all side-chain hydrogen resonances of residue are visibible in 3D-HCCH-TOCSY. 2D-[$^{1}H/^{13}C$]HSQC spectrum which is correlation between ^{1}H and ^{13}C was also used for verification of side-chain assignment. From the backbone assignment of BldD-CTD, chemical shifts of $^{13}C^{\alpha}(i)$, $^{1}H^{\alpha}(i)$, $^{13}C^{\beta}(i)$ were used for peak clustering of 3D-HCCH-COSY, HCCH-TOCY and 2D-[$^{1}H/^{13}C$]HSQC as shown in Fig. 6. Using the difference of correlation between HCCH-TOCSY and HCCH-COSY, 98.3 % of side-chain including chemical shifts of $^{13}C^{\alpha}$, $^{1}H^{\alpha}$, $^{13}C^{\beta}$ were assigned (Fig. 7).

5. NOESY cross-peak assignment of BldD-CTD

The NOESY allows correlating nuclei through space (distance smaller than 5Å). By measuring cross-peak intensity or volume, distance information can be

acquired (Wüthrich K, 1986). To avoid overlap problem of NOE cross-peak measured from BldD-CTD, 3D-[^{15}N]-edited NOESY and 3D-[^{13}C]-edited NOESY spectra were used for NOE assignment (Marion *et al*., 1989). Based on assigned chemical shifts, NOE cross-peaks in each diagonal peak were assigned (Fig. 8). As information of distances is already known for intra-residues, intensities of intra-residual cross-peaks are used as distance standards (Wüthrich K, 1986). To calculate 3D protein structure, the set of lower and upper distance limits among protons is required. As the sum of van der Waals radii for two protons (about 1.8 Å) used as a lower distance limit in general, upper distance limits are critical for determination of protein structure. From these, each upper limit distance for strong, medium, and weak cross-peak was determined to 2.5-3.0 Å, 4.0 Å, and 5.0 Å. Among them, long-range NOEs, which are correlations between atoms separated by five or more residues along the protein sequence, provide important 3D-structural information for protein (Wüthrich K, 1986). The upper distance limits for long-range NOEs of BldD-CTD listed in Table 2.

6. 3D-Structure determination using CYANA

Using the upper distance limits from assignment of NOE cross-peaks, initial structure of BldD-CTD calculated by dihedral torsion angle dynamics

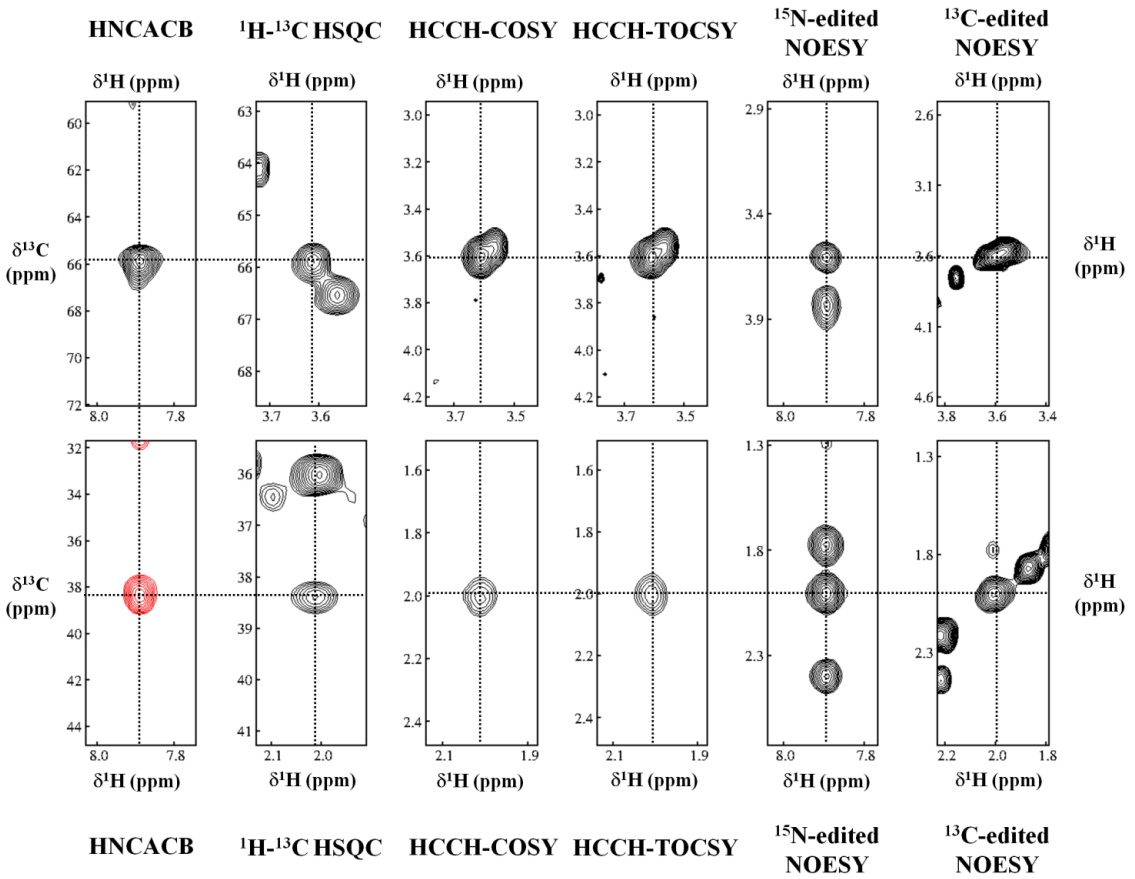

HNCACB
¹H-¹³C HSQC
HCCH-COSY
HCCH-TOCSY
¹⁵N-edited NOESY
¹³C-edited NOESY
δ¹H (ppm)
δ¹³C (ppm)
δ¹H (ppm)
δ¹³C (ppm)
δ¹H (ppm)
HNCACB
¹H-¹³C HSQC
HCCH-COSY
HCCH-TOCSY
¹⁵N-edited NOESY
¹³C-edited NOESY

Fig. 6. Verification of a peak cluster of Ile135 for side-chain assignment. The 6 peaks for $^{13}C^{\alpha}$ (upper line) and 6 peaks for $^{13}C^{\beta}$ (bottom line) are combined into a cluster. The X-axis of the individual spectrum represents the ^{1}H chemical shifts. The Y-axis of the HNCACB and 2D-[$^{1}H/^{13}C$]HSQC spectrum represent ^{13}C chemical shifts, and those of the other spectra represents $^{1}H^{\alpha}$ (upper line) or $^{1}H^{\beta}$ (bottom line) chemical shifts. 3D-HNCACB and 3D-[^{15}N]-edited NOESY have the same $^{15}N^{H}$-edited slide, and other spectra except 2D-[$^{1}H/^{13}C$]HSQC have the same $^{13}C^{\alpha}$- (upper line) or $^{13}C^{\beta}$- (bottom line) edited slide.

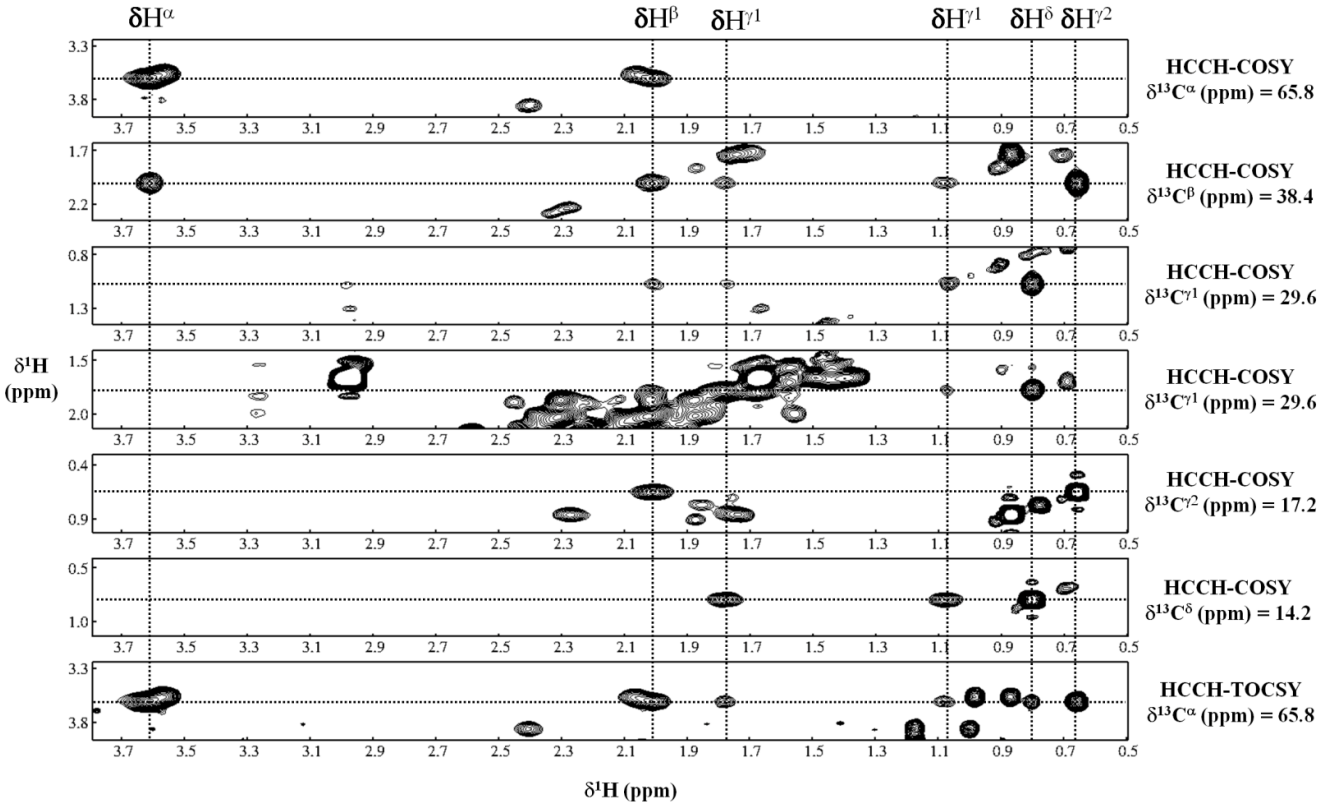

δH^α
δH^β
δH^γ1
δH^γ1
δH^δ
δH^γ2
HCCH-COSY
δ13Cα (ppm) = 65.8
HCCH-COSY
δ13Cβ (ppm) = 38.4
HCCH-COSY
δ13Cγ1 (ppm) = 29.6
HCCH-COSY
δ13Cγ1 (ppm) = 29.6
HCCH-COSY
δ13Cγ2 (ppm) = 17.2
HCCH-COSY
δ13Cδ (ppm) = 14.2
HCCH-TOCSY
δ13Cα (ppm) = 65.8
δ1H (ppm)
δ1H (ppm)

Fig. 7. Side-chain assignment of Ile35 using HCCH-COSY and HCCH-TOCSY. Each spectrum is labeled on its right side. ^{13}C chemical shifts for editing spectrum are also labeled on its right side. Above vertical dotted lines ^{1}H chemical shifts correlated among spectra are labeled. Horizontal dotted lines represent the ^{1}H chemical shifts correlated in each spectrum.

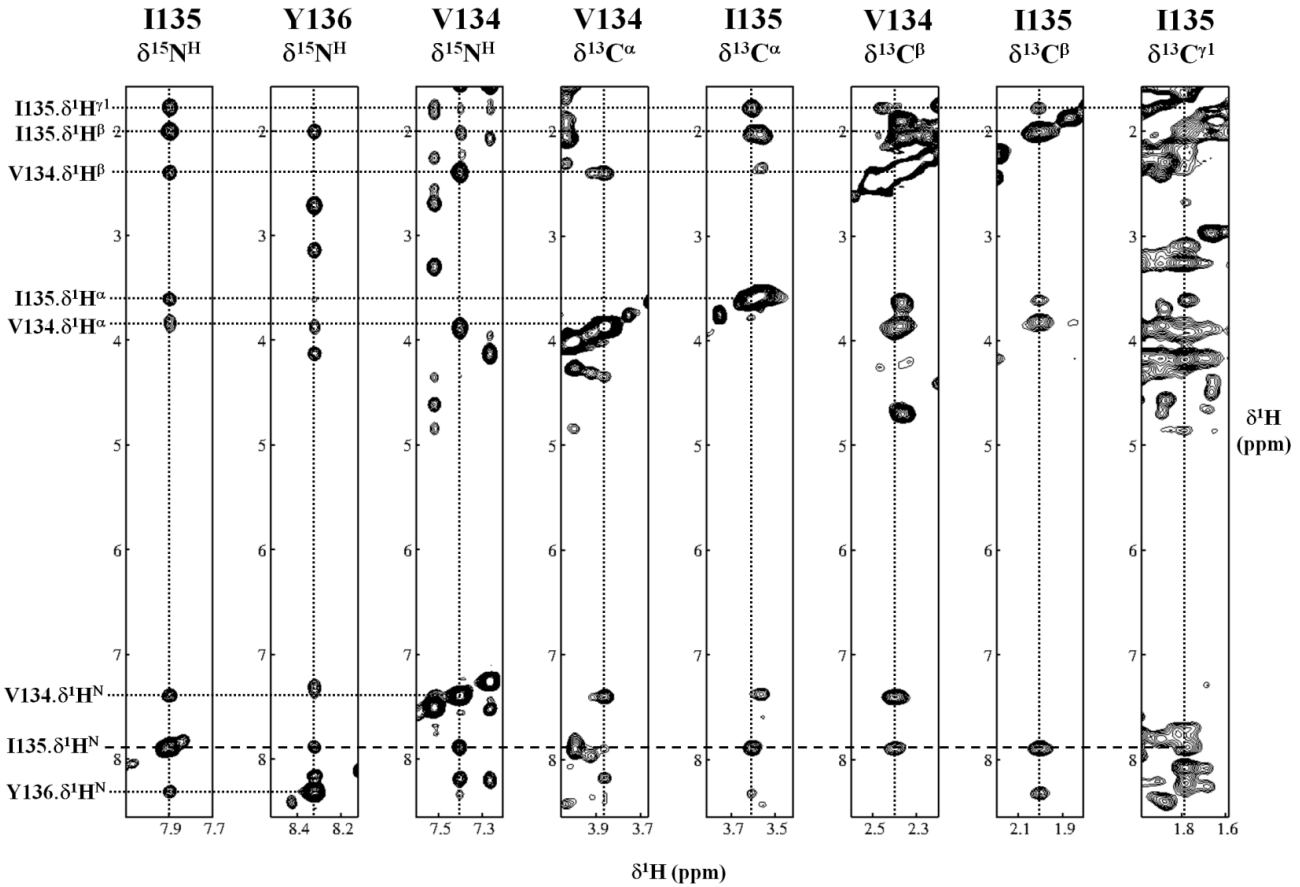

I135 δ15NH
Y136 δ15NH
V134 δ15NH
V134 δ13Cα
I135 δ13Cα
V134 δ13Cβ
I135 δ13Cβ
I135 δ13Cγ1
I135.δ1Hγ1
I135.δ1Hβ
V134.δ1Hβ
I135.δ1Hα
V134.δ1Hα
V134.δ1HN
I135.δ1HN
Y136.δ1HN
δ1H (ppm)
δ1H (ppm)

Fig. 8. NOE cross-peaks assignment of Ile135 on 3D-[^{15}N]-edited NOESY. Above vertical dotted lines in each spectrum, corresponding residue and ^{1}H chemical shift are labeled. Among the 8 strip plots, 3 plots on the left side are 3D-[^{15}N]-edited NOESY and other plots are 3D-[^{13}C]-edited NOESY. From the correlation among spectra, assignments of NOE cross-peaks for Ile135 are labeled on the left side of spectrum corresponding to Ile135.

Table 2. Upper distance limits for long-range NOEs. The nomenclature of atoms in amino acid residues follows the IUPAC recommendations. 'A', 'B', 'G', 'D', and 'E' represent 'α', 'β', 'γ', 'δ', and 'ε', respectively. 'Q' represents pseudoatom (Markley et al, 1998). 'Upl' represents upper distance limit.

Residue	Atom	Residue	Atom	Upl (Å)	Residue	Atom	Residue	Atom	Upl (Å)
83 PRO	HB2	157 ALA	HA	5.5	95 VAL	QG2	100 ALA	QB	5.5
83 PRO	QG	124 ILE	H	6	95 VAL	QG1	100 ALA	QB	6
84 LYS	QG	120 LYS	HA	6.5	95 VAL	QG2	150 TRP	HE1	6.5
86 VAL	QG2	121 VAL	HA	6	99 LYS	HD3	150 TRP	HE1	5.5
86 VAL	HA	123 SER	HA	5	99 LYS	HD2	150 TRP	HE1	5.5
86 VAL	QG1	123 SER	HA	6	103 LEU	QD2	108 ALA	H	6.5
86 VAL	QG1	123 SER	QB	6	103 LEU	HA	132 LEU	QD1	6
86 VAL	QG2	123 SER	HA	5	103 LEU	HB3	136 TYR	HE1	5.5
86 VAL	HA	124 ILE	H	5	103 LEU	HB3	136 TYR	HE2	5.5
87 LEU	H	122 LEU	H	5	103 LEU	HB2	136 TYR	HE1	5.5
88 ASP	H	152 VAL	HA	5	103 LEU	HB2	136 TYR	HE2	5.5
88 ASP	H	152 VAL	QG2	5.5	103 LEU	QD1	136 TYR	HE1	6.5
88 ASP	HB2	152 VAL	HA	5	103 LEU	QD1	136 TYR	HE2	6.5
88 ASP	HB3	152 VAL	HA	4	103 LEU	QD2	136 TYR	HE1	6.5
89 LEU	QD2	107 ALA	QB	5	103 LEU	QD2	136 TYR	HE2	6.5
89 LEU	QD1	111 GLN	HE22	6.5	103 LEU	QD1	147 LEU	QD1	6
89 LEU	QD1	111 GLN	HE21	6	103 LEU	QD1	147 LEU	QD2	6
89 LEU	H	121 VAL	QG1	6	103 LEU	QD1	150 TRP	HZ3	6.5
89 LEU	H	121 VAL	HA	4	103 LEU	QD2	150 TRP	HZ3	6.5
89 LEU	QD1	121 VAL	HA	6.5	106 TYR	HB3	132 LEU	QD1	6
89 LEU	QD2	121 VAL	HA	6.5	106 TYR	HB2	132 LEU	QD1	6
89 LEU	QD1	122 LEU	H	6	106 TYR	QE	128 ASP	HB2	6.5
89 LEU	QD2	122 LEU	H	6.5	107 ALA	H	132 LEU	QD1	6.5
89 LEU	HG	121 VAL	HA	5	107 ALA	HA	132 LEU	QD1	6
92 LEU	QD1	104 GLN	H	6	111 GLN	HA	116 ASP	HB2	5.5
92 LEU	QD1	104 GLN	HE21	6	111 GLN	HA	116 ASP	HB3	5.5
92 LEU	QD1	104 GLN	HE22	6	111 GLN	HE21	117 TYR	HA	6
93 ALA	HA	104 GLN	HE22	5	111 GLN	HE22	117 TYR	HA	6
93 ALA	HA	104 GLN	HE21	6	111 GLN	HE22	122 LEU	QD1	6.5
95 VAL	H	100 ALA	QB	6	111 GLN	HG2	122 LEU	QD1	6
95 VAL	HB	100 ALA	QB	5	111 GLN	HG3	122 LEU	QD1	6
95 VAL	QG1	100 ALA	H	6	133 ALA	H	143 LEU	QD2	6
95 VAL	QG2	100 ALA	H	6	137 ASP	H	143 LEU	QD1	5.5
95 VAL	QG2	100 ALA	HA	6.5	148 ILE	QD1	158 ARG	H	6

algorithm DYANA embedded in CYANA 2.1 program (Herrmann *et al*., 2002). The initial ensemble structure of BldD-CTD showed average backbone RMSD of 1.73 ± 0.48 Å, which fulfilled the quality required for reliable structure (Güntert, 2004) (Fig. 9A). As shown in Fig. 9B, representative initial structure of BldD-CTD had four α-helices and a weak helical region (denoted as η1).

To obtain more NOE assignments which were not identified by manual assignment, and higher resolution of 3D-structure, CANDID module for automated NOE assignment was used for calculation of structure. As the CANDID module uses algorithms of ambiguous peak assignment and network anchoring during seven cycles which uses NOE assignment and calculated structure of the previous cycle as templates, 3D-structure of high quality can be calculated (Güntert, 2004). As expected that, fine 3D-structure of BldD-CTD (Average backbone RMSD = 0.60 ± 0.11 Å) was calculated from combination of CANDID and DYANA (Fig. 10A). It also had four α-helices and a weak helical region η1, which was identical to initial structure from manual NOE assignment (Fig. 10B). In addition, although RMSD between two representative structures over 73 atoms of C^{α} (P83-R114 and K120-A160) was 2.330 Å, as ensemble structure from manual assignment had higher RMSD than that from automated assignment and secondary structure regions were identical, two

structures from different methods are almost same (Fig. 11). For reliable structure calculation from automated NOE assignment, there are six criteria as follows (Güntert, 2004):

1. Average CYANA target function value of cycle 1 below 250 $Å^2$.
2. Average final CYANA target function value below 10 $Å^2$.
3. Less than 20% unassigned NOEs.
4. Less than 20% discarded long-range NOEs.
5. RMSD value in cycle 1 below 3 Å.
6. RMSD between the mean structures of the first and last cycle below 3 Å.

As shown in Table 3, all criteria except criterion 4 were fulfilled for automated structure calculation of BldD-CTD. The percentage of discarded long-range NOEs cannot be calculated readily outside the CYANA program because it requires knowledge of the possible assignments. In this case, an overall percentage of unused cross-peaks less than 15% can be used as an alternative criterion (Güntert, 2004). In the cycle 7, overall proportion of unassigned cross-peaks was only 8.6 %, which satisfied alternative criterion 4. The β-sheet predicted from CSI method and TALOS (K84-L87 and V121-I124) was not clearly seen in both calculated structures (Fig. 9 and 10). But, because long-range NOEs are detected in these regions (Table 2) and they are not

refined structures, it is so reasonable that a β-sheet exists in these regions.

7. Refinement of the structure using CNS

For improving efficiency of three dimensional protein structure calcuation, simplification of the nonbonded interactions was applied to the process of calculation. This causes unreal treatment of electrostatic and van der Waals (vdW) interactions, followed by unsatisfactory quality indices from validation programs. Therefore, refinement process is essential for bringing the protein structures closer to physical reality (Linge *et al*., 2003).

To refine the structure of the BldD-CTD, CNS program (Brünger *et al*., 1998) using RECOORD scripts (Nederveen *et al*., 2005) were used for recalculation of the BldD-CTD structure. As a result, final 20 structures of the BldD-CTD were calculated (Fig. 12) and validated using PSVS 1.5 (Bhattacharya *et al*., 2007) software suites (as summarized in Table 4).

8. Overall structure of BldD-CTD

The solution NMR structure of BldD-CTD (Fig. 12 and 13) consists of four α-helices, a two-stranded anti-parallel β-sheet and a C-terminal weak helical region (designated as η1) with topology of βααβααη. Helices α1 (L89-

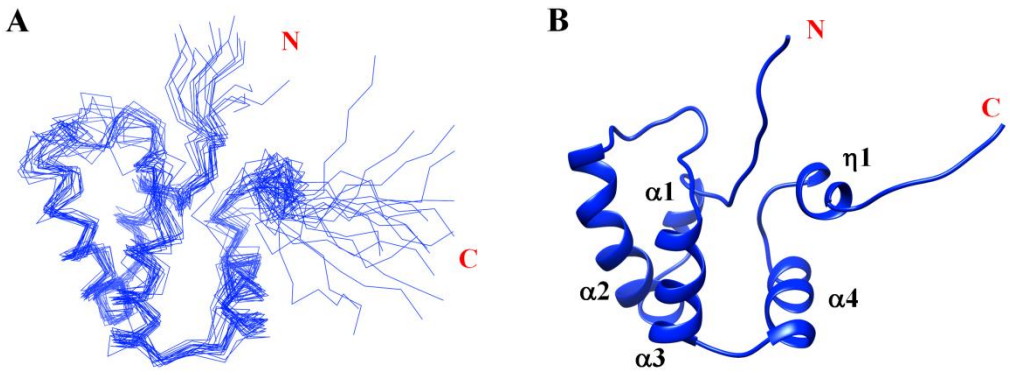


Fig. 9. Initial BldD-CTD structure calculated from manual assignment of NOE cross-peaks. (A) 20 structures of BldD-CTD calculated from manual assignment of NOE cross-peaks are superimposed over equivalent C^{α} atoms in the ordered regions (P83-R114 and K120-A160). (B) Cartoon representation for representative structure of (A). Secondary structure elements are labeled. 'N' and 'C' represent N-terminus and C-terminus, respectively.

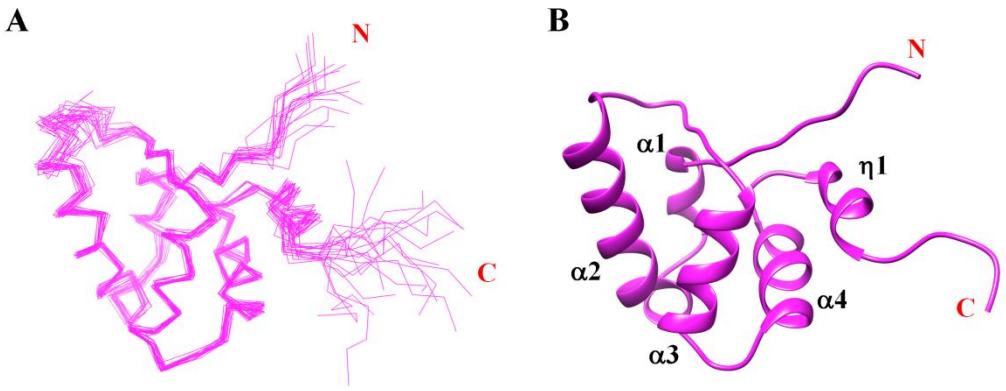


Fig. 10. BldD-CTD structure calculated from automated assignment of NOE cross-peaks. (A) 20 structures of BldD-CTD calculated from automated assignment of NOE cross-peaks are superimposed over equivalent C^{α} atoms in the ordered regions (P83-R114 and K120-A160). (B) Cartoon representation for representative structure of (A). Secondary structure elements are labeled. 'N' and 'C' represent N-terminus and C-terminus, respectively.

Table 3. Statistics of automated BldD-CTD structure calculation

Cycle	1	2	3	4	5	6	7	
Peaks:								
selected	3060	3060	3060	3060	3060	3060	3060	
assigned	2805	2845	2833	2833	2834	2830	2826	
unassigned	255	2215	227	227	226	230	234	
with diagonal assignment	210	212	212	212	212	212	212	
Cross peaks:								
with off-diagonal assignment	2595	2633	2621	2621	2622	2618	2614	
with unique assignment	985	1652	1857	1956	2088	2151	2172	
with short-range assignment [i-j] <=1	2023	2002	1978	1969	1950	1933	1934	
with medium range assignment 1<[i-j]<5	387	408	388	389	387	394	388	
with long-range assignment [i-j] >=5	185	223	255	263	285	291	292	
Upper distance limits:								
total	1718	1604	1523	1504	1468	1447	1467	1510
short-range [i-j] <=1	1148	1025	953	932	891	866	788	820
medium-range 1<[i-j]<5	523	511	341	337	329	332	359	363
long range [i-j] >=5	47	68	229	235	248	249	320	327
Average assignments/constraint	4.21	2.23	1.46	1.42	1.32	1.28	1.00	1.00
Average target function value	2.60	1.93	4.29	0.59	0.42	0.35	0.36	0.24
RMSD (residues 5-83):								
Average backbone RMSD to mean	1.45	0.95	0.70	0.75	0.70	0.72	0.61	0.60
Average heavy atom RMSD to mean	2.00	1.52	1.25	1.25	1.23	1.25	1.16	1.13

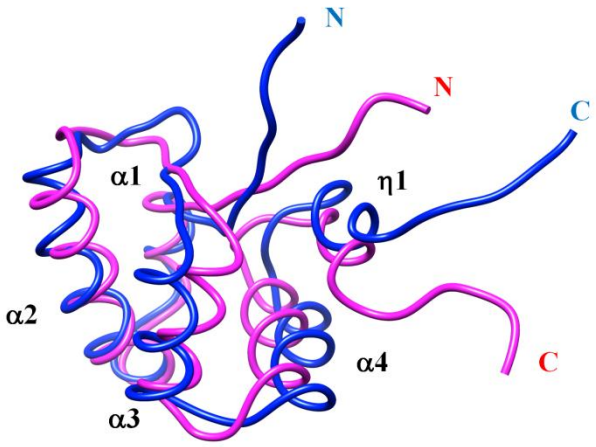


Fig. 11. Superimposed representative structures calculated from manual (blue) and automated (magenta) assignment of NOE cross-peaks. The representative structures from both ensemble structures with manual and automated assignment are superimposed over equivalent C^{α} atoms in the ordered regions (P83-R114 and K120-A160). Secondary structure elements are labeled. 'N' and 'C' represent N-terminus and C-terminus, respectively.

L92), α2 (A97-R114), α3 (L129-136) and α4 (P140-W150) form the core frame and a β-sheet composed of strand β1 (L85-D88) and β2 (V121-I124) is packed against the one side of helix bundle (Fig. 13). The overall conformation of BldD-CTD is stabilized mainly through the interaction of interior or semi-interior hydrophobic residues (L85 and L87 from β1, L89 and L92 from α1, L103 and Y106 from α2, L122 and I124 from β2, L129, L132, I135, and Y136 from α3, L143, L147, and W150 from α4, and V152 and L153 from the α4-η1 loop) maintains the main framework (Fig. 14). As seen in Fig. 12, all regions of BldD-CTD are well-ordered except for its N-terminal linker for BldD-NTD, a β-turn between helix α2 and strand β2, and a C-terminal tail. The helix η1 (A155-A161) adjacent to C-terminus shows a weakly ordered but dynamic property in its ensemble structure (Fig. 12). It appears as a hybrid of a turn of 3_{10}-helix (A155-A157) and a turn of α-helix (R158-A161) (Fig. 15).

To verify whether this helix is an erroneous structure calculated from wrong assignment, the NOE cross-peaks in this region were analyzed. As shown in Fig. 16, the region from A155 to A157 shows NOEs of $d_{\alpha N}(i,i+2)$ cross-peaks, which are typical NOEs for 3_{10}-helix,. But, NOEs of $d_{\alpha N}(i,i+4)$ for α-helix were not detected in the region from R158 to A161. Since NOEs of $d_{\alpha N}(i,i+4)$ for α-helix is very weak (Wüthrich K, 1986) and a turn of α-helix in

helix η1 is close to severely dynamic C-terminus (Fig. 12), the correlations for α-helical region of η1 might be undetectable. Although the NOEs for this region were not detected sequentially, residue R159 positioned in α-helical region of η1 showed a correlation of $d_{\alpha N}(i,i+3)$ and a strong correlations of $d_{\alpha\beta}(i,i+3)$, both of which are NOEs of α-helix (Fig. 17). Thus, it can be mentioned that this region (R158-A161) has a α-helical property.

In the [^{1}H,^{15}N]-heteronuclear NOE experiments, the determined NOE values in the η1 region were relatively low to those in other helices but higher than other disordered regions (Fig. 18), which verified ordered but dynamic property of η1. In spite of dynamic property of η1, its spatial location is maintained by loose hydrophobic packing composed of A155 and A157 from the η1, L85 from the β1, I148 from the α4, and L153 from the α4-η1 loop (Fig. 14). Thus, η1 is seemed to slightly restrict dynamic feature of C-terminal region in the BldD-CTD structure.

9. Structural comparison between BldD-NTD and BldD-CTD

BldD-NTD adopts a compact globular structure composed of four α-helices, which is packed through hydrophobic interaction among helices. Helix α2 and α3 constitute a canonical helix-turn-helix fold and helix α1 forms

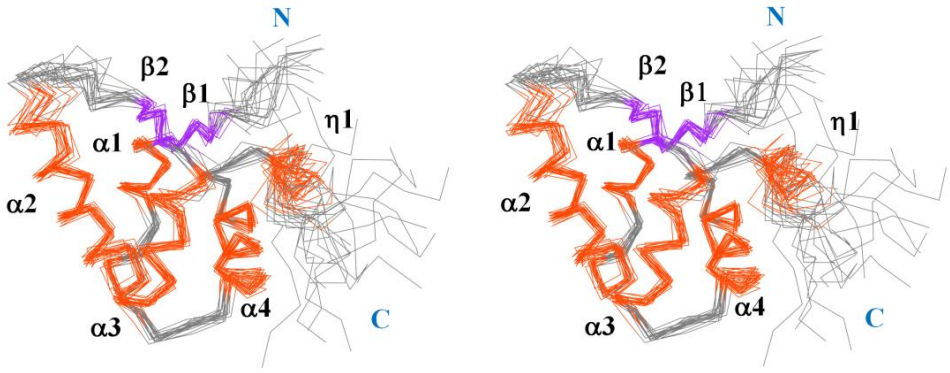


Fig. 12. Stereoview of the BldD-CTD ensemble structure from refinement process. 20 final structures of BldD-CTD are superimposed over equivalent C^{α} atoms in the ordered regions (P83-R114 and K120-A160). Secondary structure elements are labeled and colored orange for α- and η-helices and purple for β-strands.

Table 4. Summary of NMR and Structural Statistics for the final 20 conformers of the BldD-CTD[a]

Completeness of resonance assignments	
Backbone (%)	97.6
Side chain (%)	98.3
Restraint statistics	
Distance Restraints	
Total	1382
Intra-residue (i = j)	311
Sequential (|i – j| = 1)	381
Medium range (1 < |i – j| < 5)	363
Long range (|i – j| ≥ 5)	327
Distance constraints per residue	17.3
Dihedral-angle constraints	138
ϕ	69
ψ	69
Number of restraints per residue	19
Number of long-range restraints per residue	4.1
CYANA targat function (Å^2)	0.24
Residual restraint violations	
Average number of distance violations per structure	
0.1–0.2 Å	1.15
0.2–0.5 Å	0.1
>0.5 Å	0
Average RMS distance violation/restraint (Å)	0.00
Maximum distance violation (Å)	0.21
Average number of dihedral angle violation per structure	
1-10°	1.3
>10°	0
Average RMS of dihedral angle violation/constraint	0.18
Maximum dihedral angle violation (degree)	2.6

[a]Calculated using the PSVS 1.5 program (Bhattacharya *et al*., 2007).

Table 4. (Continued)

RMSD from average coordinates (Å)[b]		
Backbone atoms	0.7	
Heavy atoms	1.2	
Procheck Ramachandran statistics		
Most favored regions (%)	96.3	
Additional allowed regions (%)	3.7	
Generously allowed regions (%)	0.0	
Disallowed regions (%)	0.0	
Global quality scores (Raw/Z-score)		
Procheck (phi-psi)	0.13	0.83
Procheck (all)	0.03	0.18
Verify3D	0.41	-0.80
ProsaII	0.80	0.62
MolProbity clash	11.16	-0.39
RPF scores[c]		
Recall	0.953	
Precision	0.871	
F-measure	0.910	
DP-score	0.791	

[b]Analyzed in the ordered [S(phi) + S(psi) > 1.8] regions (P83-R114 and K120-A160).

[c]RPF scores reflect the goodness-of-fit of the final ensemble of structures.

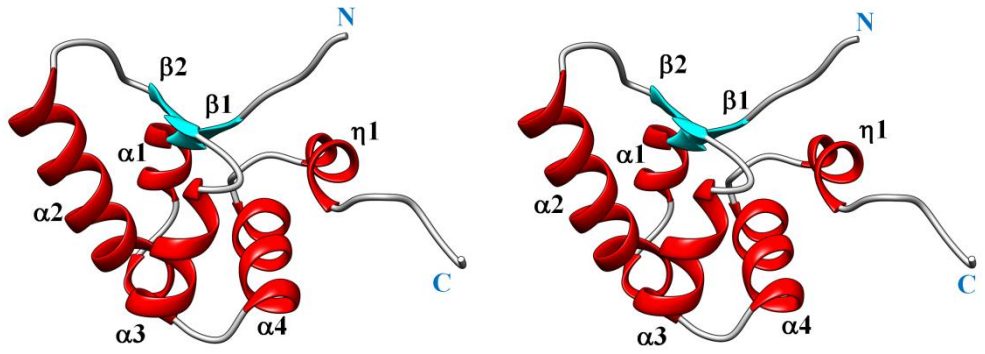


Fig. 13. Stereoview of cartoon representation for the representative BldD-CTD structure out of 20 final structures. Secondary structure elements are labeled and colored red for α- and η-helices and cyan for β-strands.

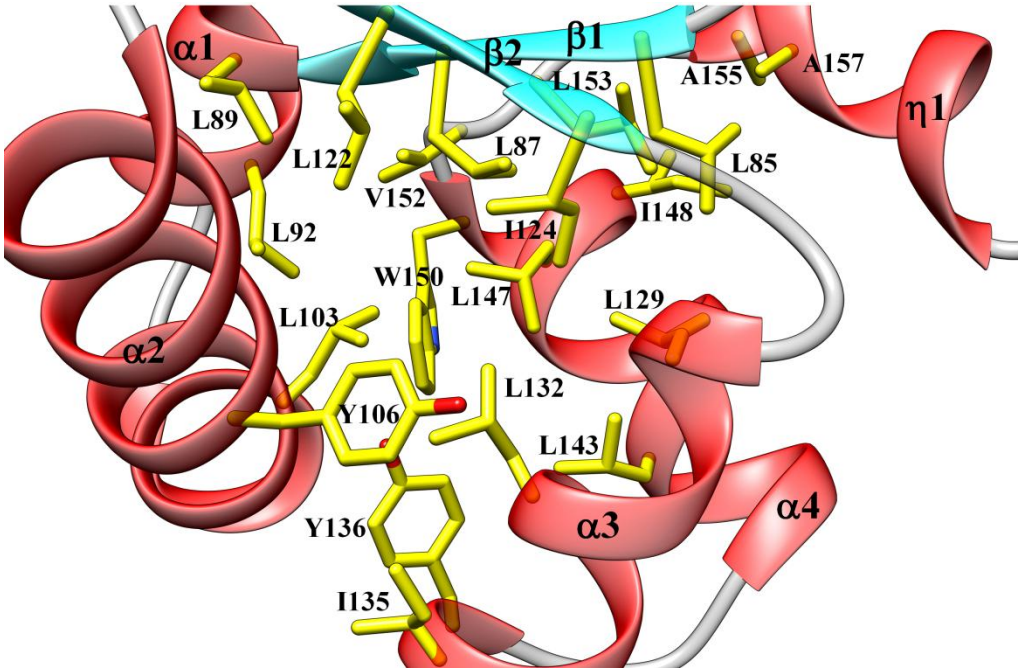


Fig. 14. Hydrophobic core formation of the BldD-CTD structure. Buried or partly buried hydrophobic side-chains are represented as yellow sticks with heteroatoms colored red for oxygen and blue for nitrogen. Ribbons of secondary structure elements are colored red for helices (α1-α4 and η1) and cyan for β-strands (β1 and β2).

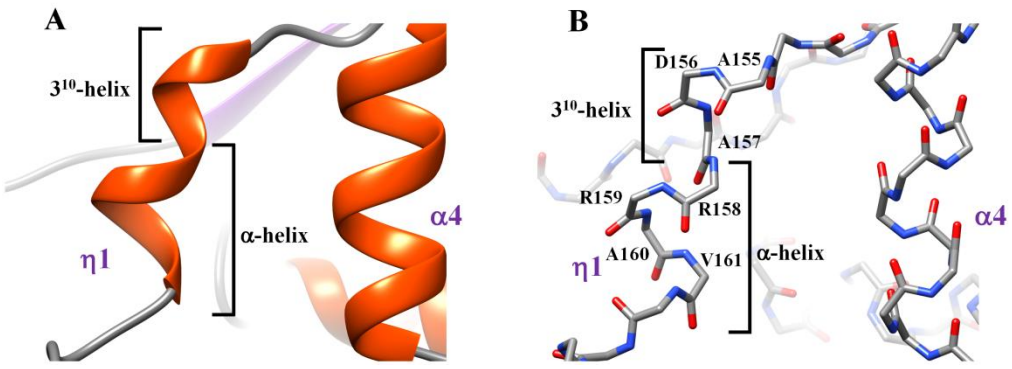


Fig. 15. Hybrid helix η1 of the BldD-CTD structure. (A) cartoon representation of helix η1. (B) Representation of backbone atoms for helix η1. Heteroatoms are colored red and blue for oxygen and nitrogen, respectively. Individual residues are also labeled. Secondary structure elements are labeled and each region for 3_{10}-helix and a-helix are indicated in (A) and (B).

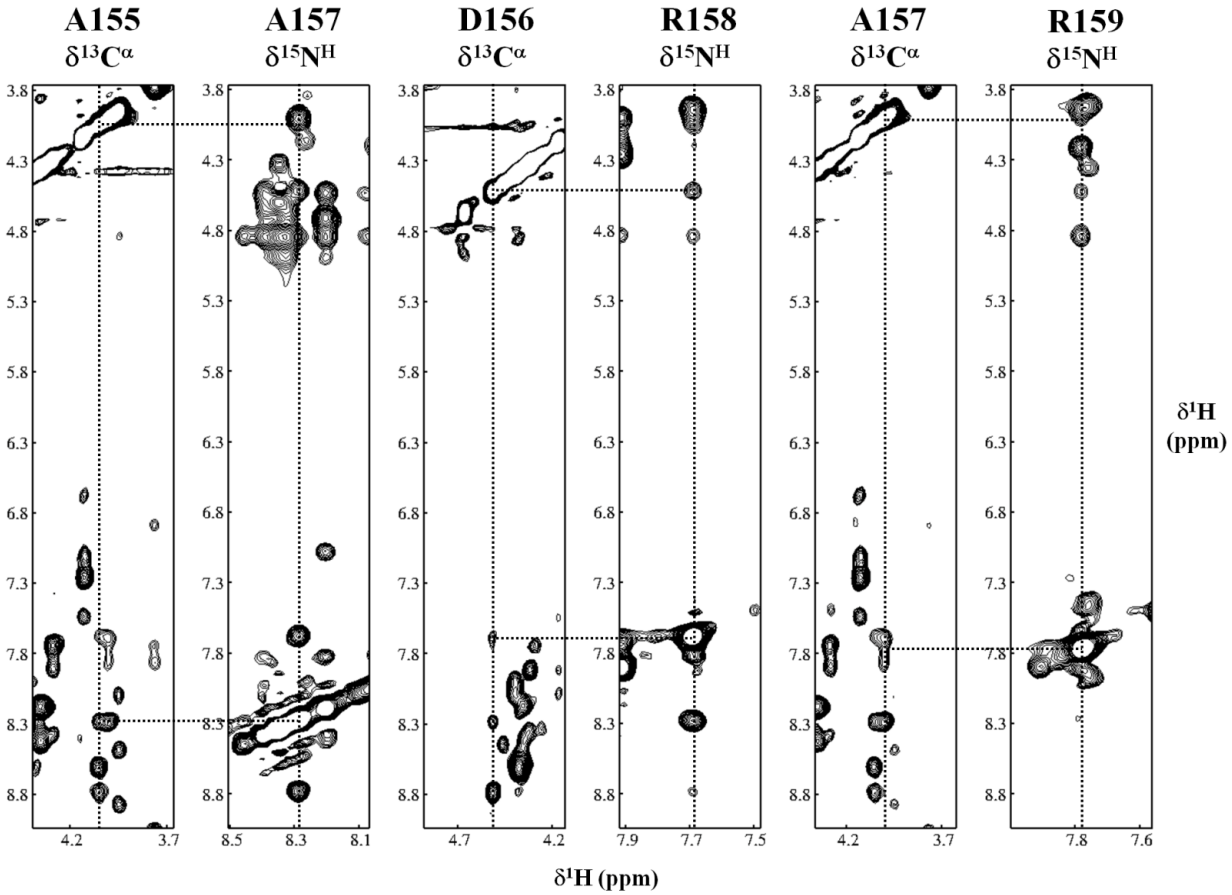

A155
δ13Cα
A157
δ15NH
D156
δ13Cα
R158
δ15NH
A157
δ13Cα
R159
δ15NH
δ1H (ppm)
δ1H (ppm)

Fig. 16. NOE correlations of 3_{10}-helix region from η1 in the BldD-CTD structure. Above vertical dotted lines in each spectrum, corresponding residue and ^{1}H chemical shift are labeled. Correlations of $d_{\alpha N}(i,i+2)$ between spectra are represented horizontal dotted lines.

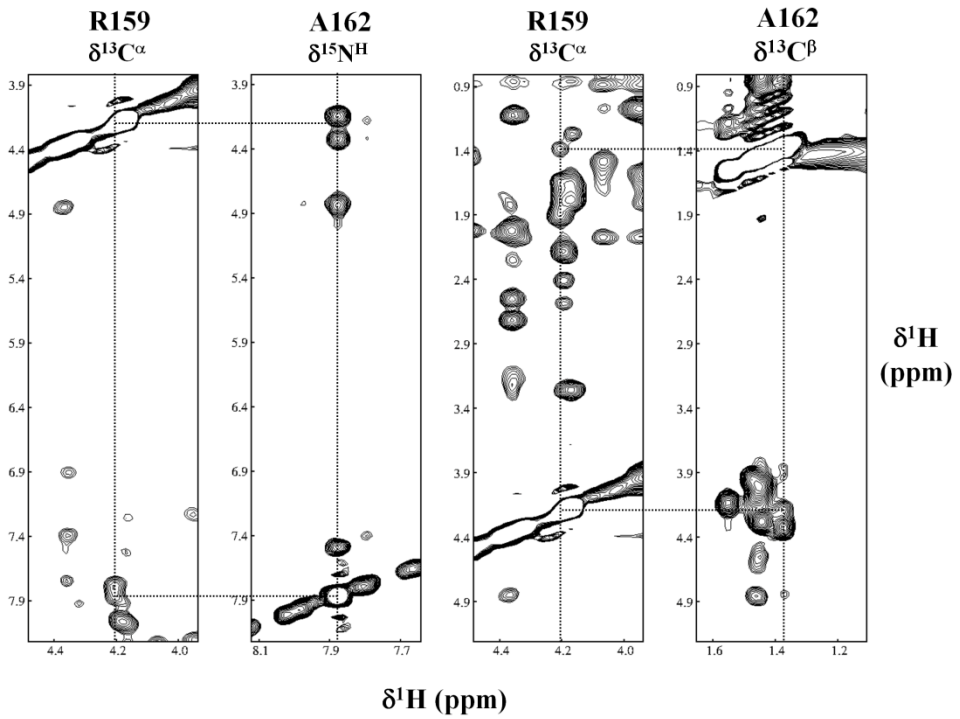


Fig. 17. NOE correlations of α-helix region from η1 in the BldD-CTD structure. Above vertical dotted lines in each spectrum, corresponding residue and ^{1}H chemical shift are labeled. Correlations of $d_{\alpha N}(i,i+3)$ or $d_{\alpha\beta}(i,i+3)$ between spectra are represented horizontal dotted lines.

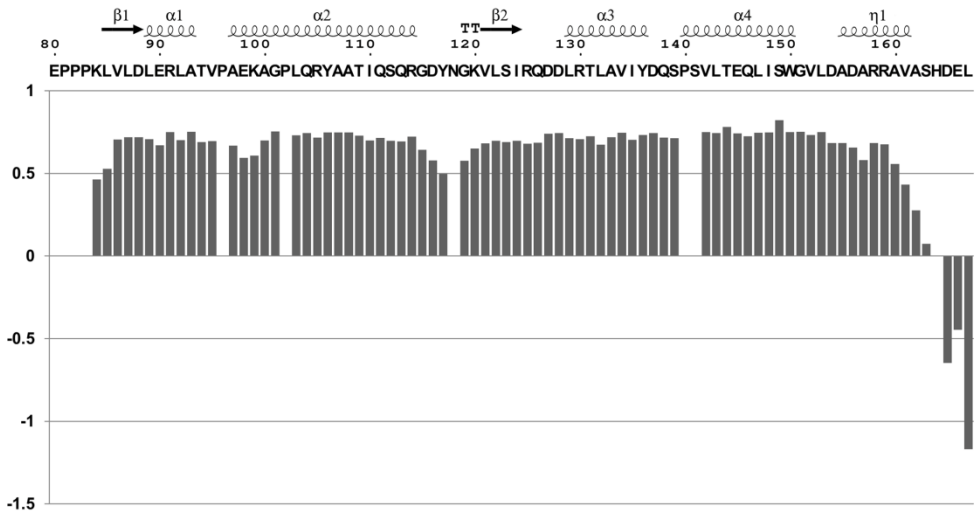


Fig. 18. [^{1}H,^{15}N]-heteronuclear NOE values of individual residues in the BldD-CTD. Secondary structure elements are illustrated over the BldD-CTD sequence, of which every tenth residue number is indicated. 'TT' represents a strict β-turn.

a structural scaffold that anchor α2 and α3. Helix α3 of BldD-NTD functions as a ‘recoginition helix’ (Kim *et al*., 2007). Likewise, BldD-CTD adopts a fold which is maintained through hydrophobic interaction of four α-helices. In the BldD-CTD structure, differently from BldD-NTD, a pair of β-sheet forming from β1 and β2 also contributes to stabilization of overall structure by which is packed against bundle of four α-helices. Interestingly, helix α3 and α4 shows a similar fold with the helix-turn-helix fold of BldD-NTD, which are anchored by helix α2 like the helix α1 of BldD-NTD (Fig. 19). However, electrostatic surface potential of BldD-CTD shows relatively negative charges (Fig. 21) while that of BldD-NTD represents strong positive charges (Fig. 20). Therefore, it might be reasonable that BldD-CTD has another function which is not related to DNA-binding.

10. Structural homology of BldD-CTD

The 3D-structure of the BldD-CTD was applied to DALI server (Holm *et al*., 2010) for searching structural homologues of the BldD-CTD. Unfortunately, searching the DALI server with BldD-CTD structure did not yield a structural homologue with a remarkable resemblance (Z-score > 4.0). However, most of the structures with relatively high Z-scores and low RMSD

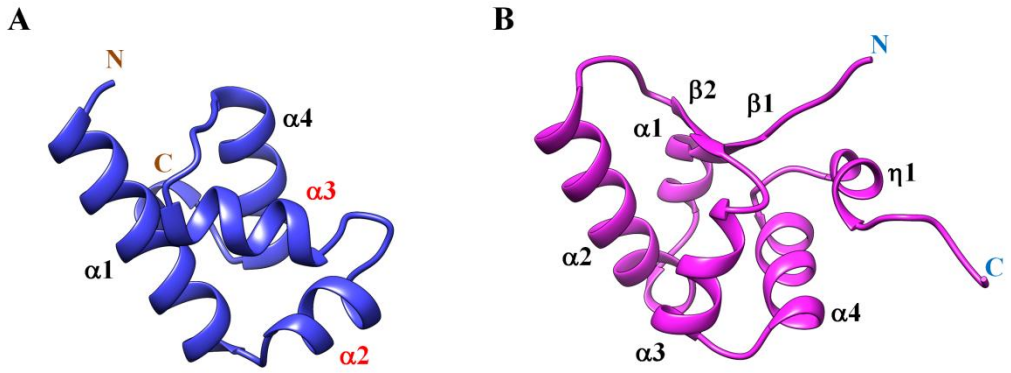


Fig. 19. Cartoon representation of (A) BldD-NTD (residues 1-79) and (B) BldD-CTD (residues 80-167). Secondary structure elements are labeled. 'N' and 'C' represent N-terminus and C-terminus, respectively. In panel (A), α2 and α3 constitutes a helix-turn-helix fold of BldD-NTD, designated as red fonts.

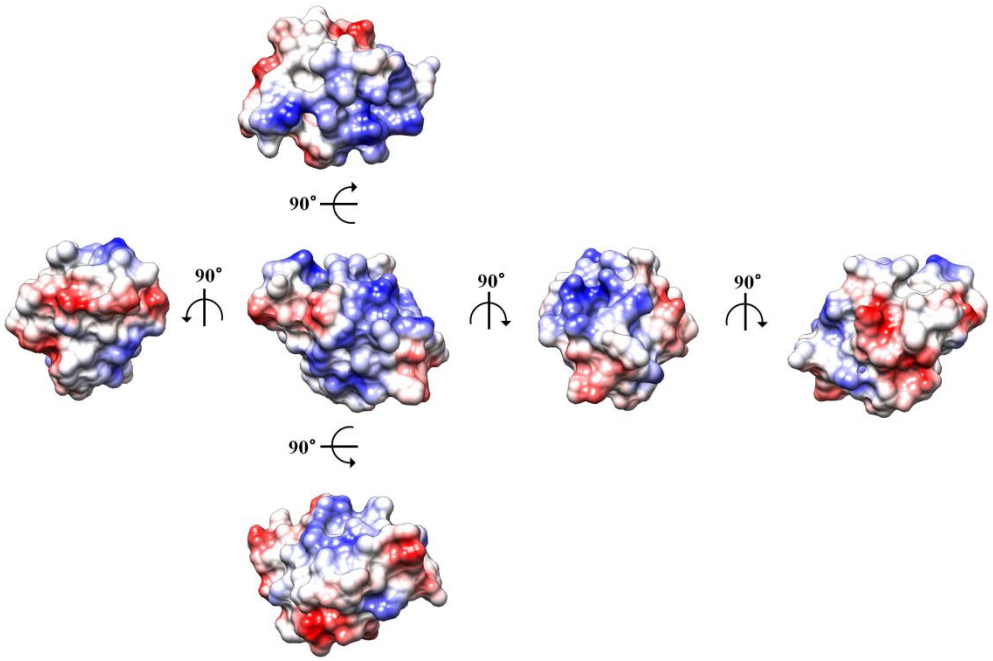


Fig. 20. Electrostatic surface potential of the BldD-NTD structure. Rendered surfaces are colored according to electrostatic surface potential at ±10 KT/e for positive (blue), negative (red) or neutral (white) charge potential. Angles and directions for rotation are labeled.

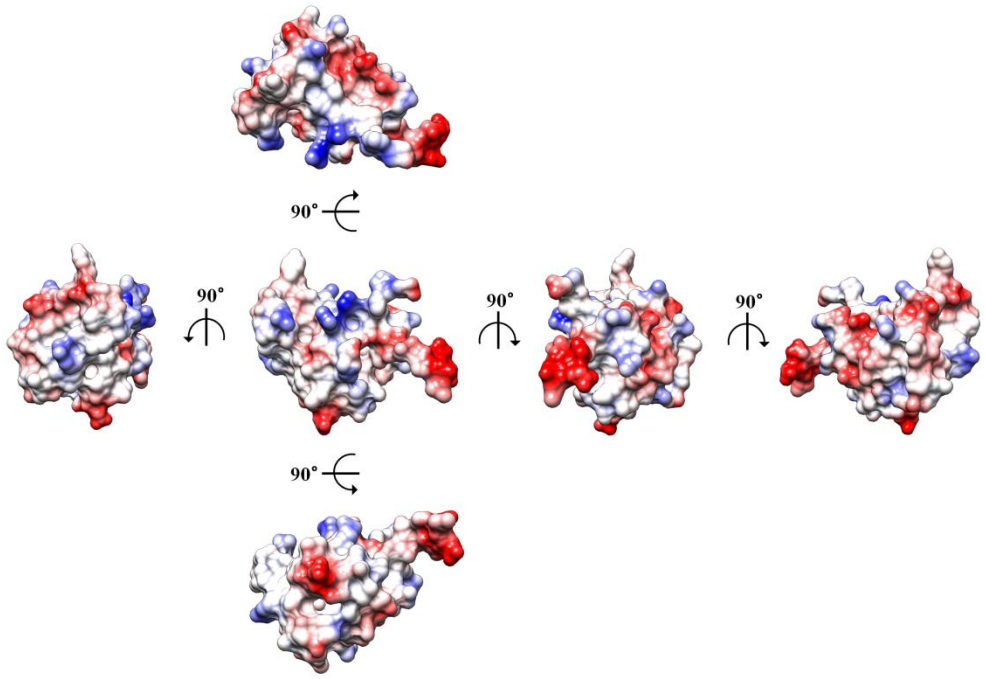


Fig. 21. Electrostatic surface potential of the BldD-CTD structure. Rendered surfaces are colored according to electrostatic surface potential at ±10 KT/e for positive (blue), negative (red) or neutral (white) charge potential. Angles and directions for rotation are labeled.

values were DNA-binding proteins containing a winged helix-turn-helix (wHTH) motif. Among them, the DNA-binding domains (DBDs) of the *Streptococcus gordonii* ScaR (PDB ID 3HRS; Z-score 2.4) and the *Thermotoga maritima* LexA (PDB ID 3K2Z; Z-score 2.6) proteins showed the most similar fold to the BldD-CTD, with an RMSD of 1.5 Å over 39 matched C^{α} atoms (Fig. 22).

ScaR-DBD and LexA-DBD (Stoll *et al*., 2009; Butala *et al*., 2009) both show a canonical winged-helix domain (WHD) with a topology of αααββ (Fig. 23). In spite of the different topology, the wHTH-like overall fold of the BldD-CTD superimposes well with those of the ScaR-DBD and LexA-DBD, particularly in the α3-α4 region (from L129 to W150), which corresponds to the helix-turn-helix (α2-α3) of the ScaR-DBD (RMSD 0.917 Å over 22 C^{α} atoms) and LexA-DBD (RMSD 1.068 Å over 22 C^{α} atoms) as shown in Fig. 24. In addition, the short two-stranded anti-parallel β-sheet (β1 and β2) of the BldD-CTD spatially compensates for the typical β-hairpin wing of the ScaR-DBD and LexA-DBD (Fig. 25).

The WHDs can interact with DNA targets thorough their wing and HTH. The wing which is the loop connecting two β-strands contacts to the minor groove of DNA and the helix-turn-helix which is formed by helix α2

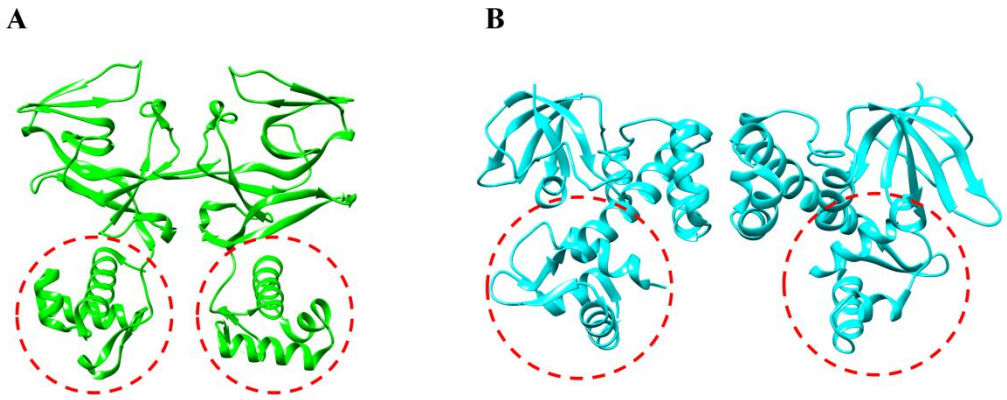


Fig. 22. Overall structures of homologues structurally similar to the BldD-CTD structure. (A) LexA from *Thermotoga maritima* (PDB ID 3K2Z) and (B) ScaR from *Streptococcus gordonii* (PDB ID 3HRS). The folds structurally similar to the BldD-CTD structure are highlighted by red-dashed circle.

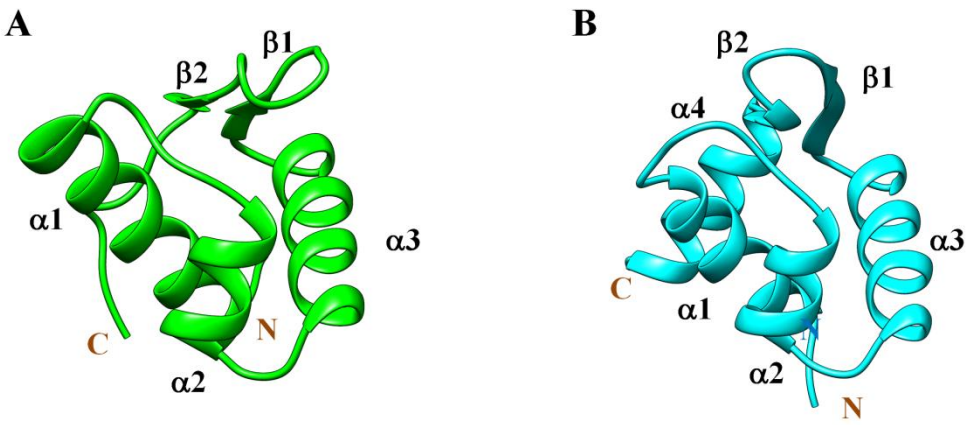


Fig. 23. Winged-helix domain of LexA (A) and ScaR (B). Secondary structure elements are labeled. ‘N’ and ‘C’ represent N-terminus and C-terminus, respectively.

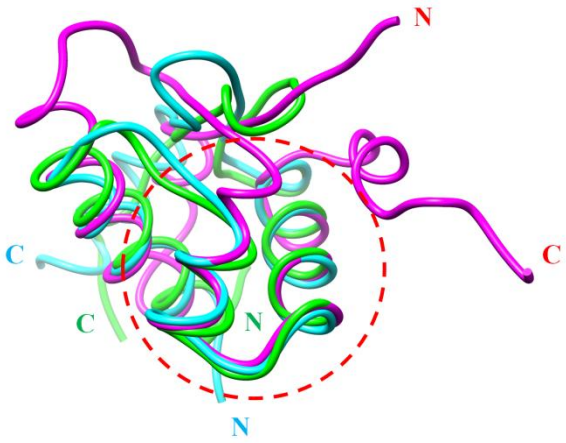


Fig. 24. Licorice style representation of superimposed BldD-CTD (magenta), LexA (green), and ScaR (cyan). The C^{α} atoms in the regions A97-Q111 and D128-G151of the BldD-CTD, E6-N20 and T23-L46 of the ScaR-WHD, and R9-G23 and S27-E50 of the LexA-WHD were superimposed. The α3-α4 region of the BldD-CTD overlays well with the corresponding regions in the other structures (red-dashed circle).

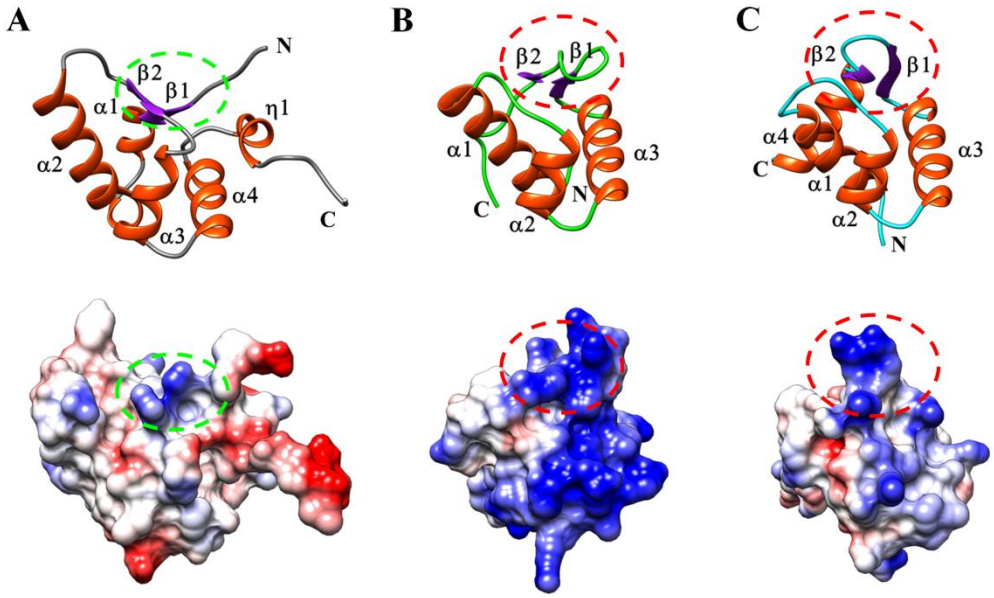


Fig. 25. Cartoon representation and electrostatic surface potential of BldD-CTD (A), LexA (B), and ScaR (C). (Top panel) Secondary structure elements are labeled. 'N' and 'C' represent N-terminus and C-terminus, respectively. (Bottom panel) Rendered surfaces are colored according to electrostatic surface potential at ±10 KT/e for positive (blue), negative (red) or neutral (white) charge potential. The β-sheet structure of BldD-CTD (A) is highlighted by green-dashed circles, and the β-hairpin structures (wing structures) of LexA and ScaR are red-dashed circles.

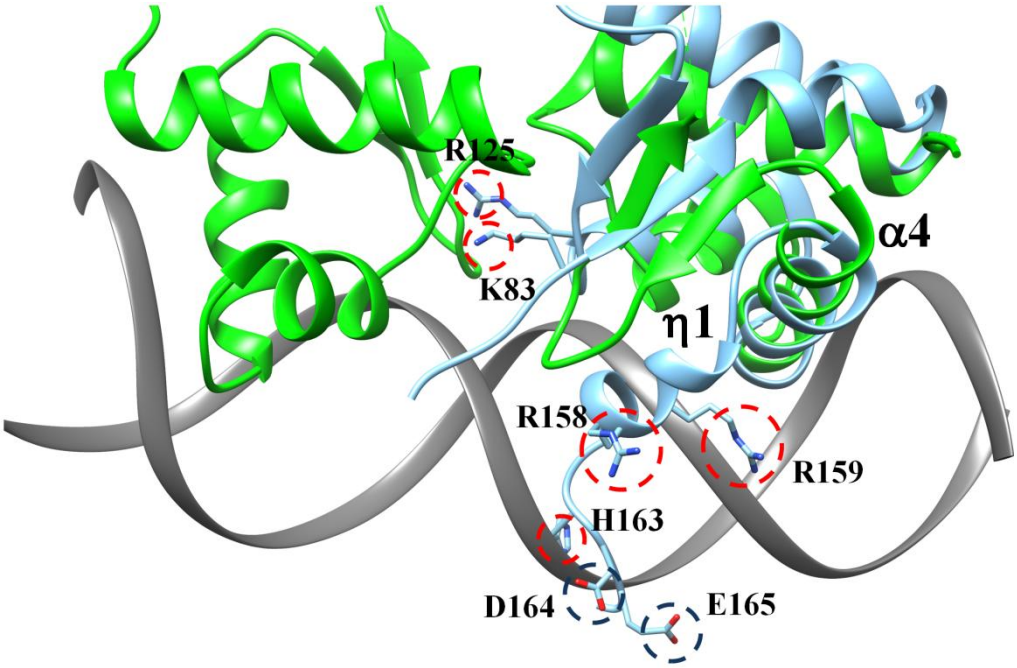


Fig. 26. Cartoon representation of superimposed BldD-CTD and LexA-DNA complex. BldD-CTD and LexA are colored cyan and green, respectively. DNA is colored gray. Positive-charged residues from wing-like structure and helix η1 from BldD-CTD are exhibited and highlighted by red-dashed circle. Negative-charged residues from C-terminus of BldD-CTD are also exhibited and highlighted by blue-dashed circle. Heteroatoms are colored red and blue for oxygen and nitrogen, respectively.

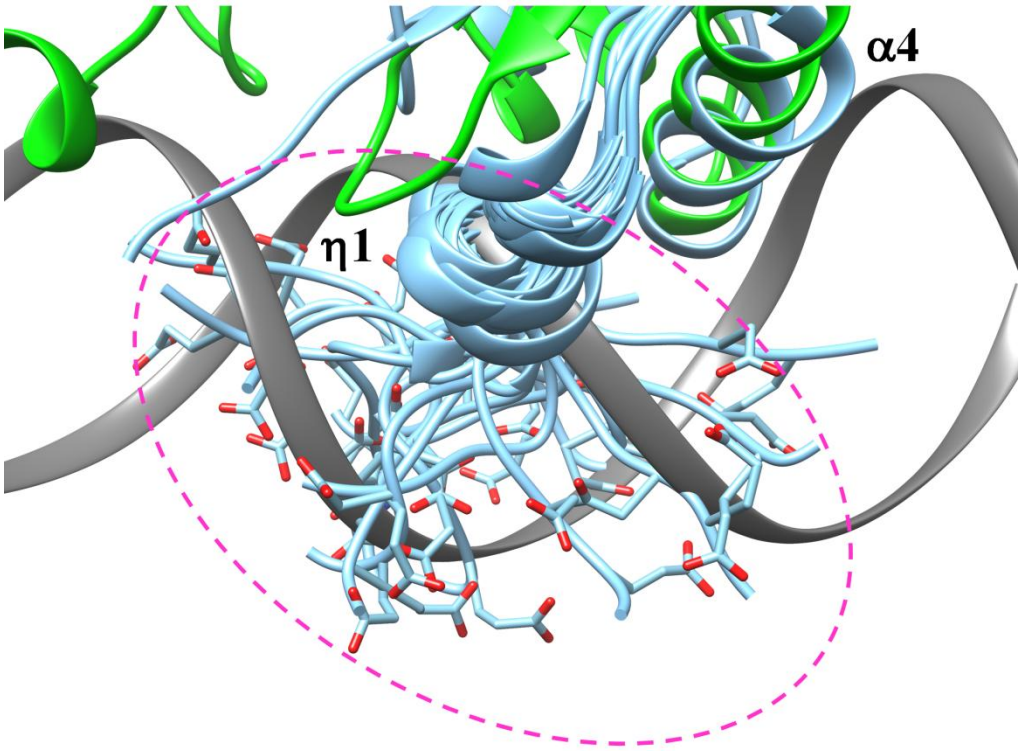


Fig. 27. Dynamic property of the BldD-CTD helix η1 and C-terminus from superimposed BldD-CTD and LexA-DNA complex. BldD-CTD and LexA are colored cyan and green, respectively. DNA is colored gray. Ensemble structure of BldD-CTD from V150 to L167 is only shown. Side-chains for D165 and E166 are shown. Heteroatom oxygen is colored red. Dynamic property of D165 and E166 is indicated by magenta-dashed line.

and α3 binds the major groove of the DNA with its helix α3 as a recognition helix (Aravind *et al*., 2005; Harami *et al*., 2013). The electrostatic surface potential analysis of the BldD-CTD shows that the face of HTH region has slightly negative charges, especially focused on its C-terminus (Fig. 27A), while its structural homologues, ScaR-DBD and LexA-DBD, have strong positive-charged surfaces on the faces of helix-turn-helix which are willing to bind the phosphate backbone of DNA (Fig. 27B and 27C). In superimposed structure of the BldD-CTD to LexA-DNA complex, hybrid helix η1 which is not existed in typical winged helix domains shows collision with DNA and negative charge residues on C-terminus of BldD-CTD (D165 and E166) might disturb DNA-binding of the BldD-CTD, although several positive charge might be positioned adjacent to DNA (Fig. 26). In addition, dynamic property of helix η1 and C-terminus could disturb the interaction between DNA and BldD-CTD (Fig. 27). Therefore, although the BldD-CTD structure is highly similar to WHD structure, it can be suggested that intact BldD-CTD has no DNA-binding activity.

11. Identification of truncated BldD

To further investigate the roles of BldD-CTD related to regulatory mechanism of BldD, the immunoblot analysis of BldD was performed

depending on morphological development of *S. coelicolor*. Interestingly, synchronized with formation of aerial mycelium, BldD appeared to be truncated in *S. coelicolor* (Fig. 28). To recognize whether BldD was truncated in N- or C-terminal region, pSET162 vector containing coding sequence of N-terminally FLAG-tagged BldD was complemented to *ΔbldD* strain. Although it could not fully complement the bald phenotype of *ΔbldD* possibly due to disturbance of attached FLAG-tag to action of BldD, delayed morphological development of this mutant strain when compared to wild type allowed to detect N-terminally FLAG-tagged BldD and its truncated form (Fig. 29A and 29B). As a result, it was confirmed that BldD was truncated in its C-terminal region.

To characterize how residues were truncated in the C-terminal region of BldD, the immunoblot analysis of 48h-cultured *S. coelicolor* was performed with various truncated forms of BldD which were heteroexpressed from pET-3a vector system in *E. coli* (Fig. 29C). It was difficult that precise site of truncation was identified, since the controls were heteroexpressed in *E. coli* and the size resolution of SDS-PAGE was limited. However, interestingly, it was observed that the truncated form corresponding to residues 1-159 of BldD had a similar size to that of truncated BldD detected in *S. coelicolor*. This truncation of BldD indicated that C-terminal region including a portion of hybrid helix η1 (A155-A161) shown in the structure of BldD-CTD (Fig. 13) was removed when

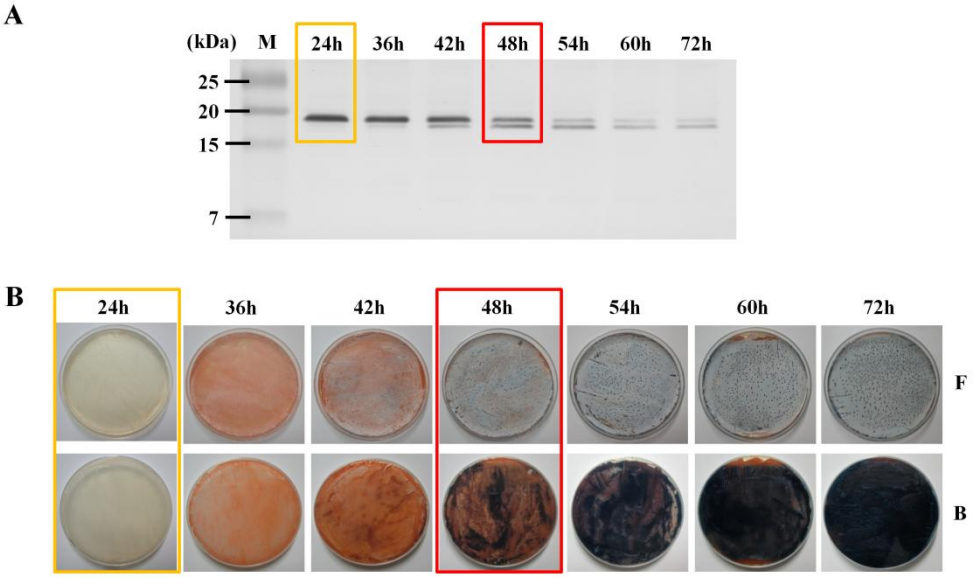


Fig. 28. Developmental stage-dependent protein patterns of BldD. (A) Immunoblot analysis of BldD in wild type. 40 μg of crude extract from each mycelia harvested at the labeled time were loaded on each lane. Cells were grown on R2YE overlaid with cellophane disc at 30 °C. (B) Phenotype of *S. coelicolor* with time-dependent morphological change. 'F' and 'B' represent front side and back side of plates, respectively.

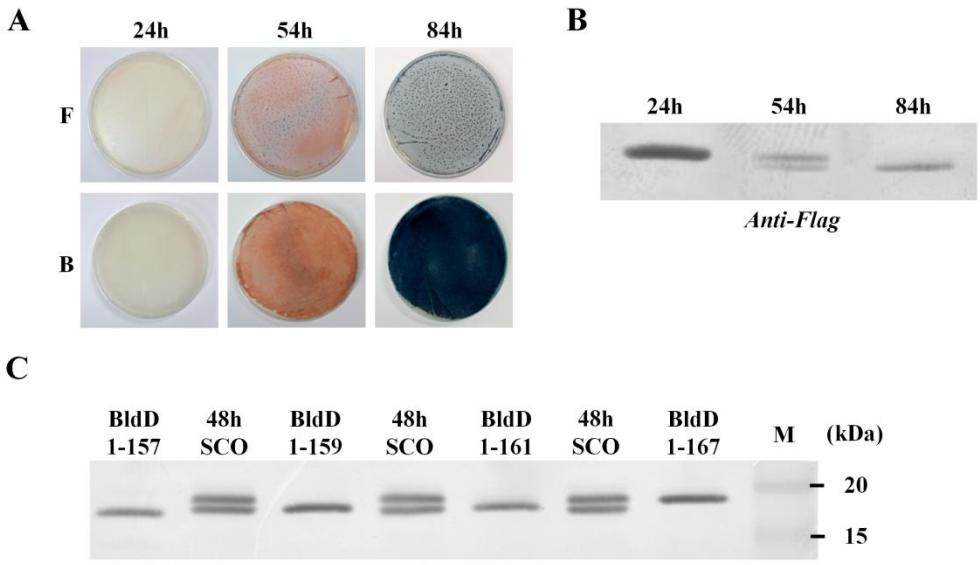


Fig. 29. Characterization of truncated BldD. (A) Phenotype of *ΔbldD* strain complemented by introduction of pSET162 derivatives containing sequence for N-terminally FLAG-tagged BldD. (B) Immunoblot analysis of N-terminally FLAG-tagged BldD. 40 μg of crude extract from each mycelia harvested at the labeled time were loaded on each lane. Cells were grown on R2YE overlaid with cellophane disc at 30 °C. (C) Verification of truncated BldD. 40 μg of crude extract from wild type cultured for 48 h were applied to immunoblot analysis. Various truncated forms of BldD heteroexpressed using pET-3a system in *E. coli* were used as standards.

S. coelicolor formed aerial mycelium.

12. Gel mobility shift assays and *in vivo* complementation experiments of truncated forms of BldD

In the stage of aerial mycelium formation of *S. coelicolor*, transcriptional level of *bldD* was considerably decreased but those of BldD-regulated genes including *bldN* and *whiG* were increased (den Hengst *et al*., 2010; Elliot *et al*., 1998; Elliot *et al*., 2001). In addition, after morphological differentiation, truncated BldD existed as small amount compared to that of substrate mycelium. These suggest that truncated BldD might act as a repressor only for regulatory region of *bldD*.

As mentioned above, in the BldD-CTD structure, the hybrid helix η1 was expected that disturb the DNA-binding of BldD-CTD (Fig. 26 and 27). In spite of negative-charged surface of BldD-CTD (Fig. 25A), it shows considerably similar structure to those of WHDs and is linked to the BldD-NTD, a strong DNA-binding domain. Thus, the truncation of helix η1 might induce BldD-CTD to bind the regulatory region of *bldD*, followed by stronger repression of truncated BldD for *bldD*. To investigate this possibility, four truncated forms of BldD-CTD and BldD were heteroexpressed in *E. coli* and purified (Fig. 30). Purified proteins were applied to gel mobility shift assays

with target DNAs of BldD. However, truncated forms of BldD-CTD did not bind any DNAs in spite of high concentration up to 100 μM (Fig. 31). On the contrary, the truncated forms of BldD showed slightly increased affinity for all DNAs (Fig. 32). As detection of truncated BldD was synchronized with morphological change of *S. coelicolor* and DNA-binding patterns were not altered in all forms of BldD, it seemed that increased DNA-binding seen in Fig. 32 was not by DNA-binding of BldD-CTD but by net charge increase due to removal of negative-charged residues, D165 and E166. In addition, *in vivo* complementation experiments for truncated forms of BldD did not show any significant difference with complementation of wild type BldD (Fig. 33). Thus, despite the unique structural features of helix η1, the role of BldD-CTD would not be related to this region.

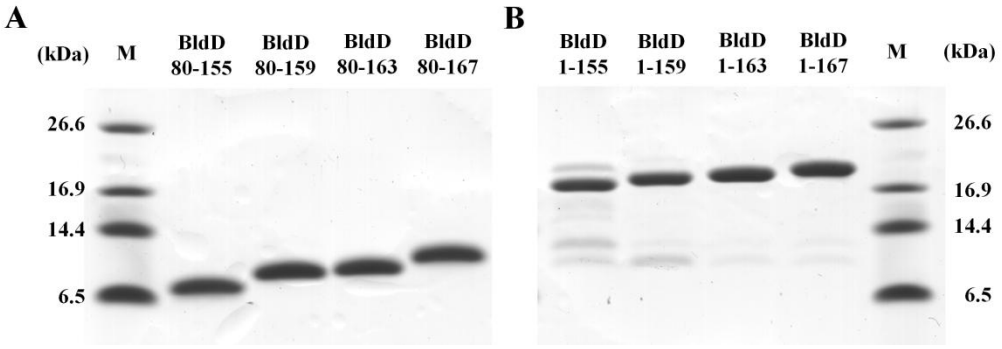


Fig. 30. Purification of various truncated forms of BldD-CTD (A) and BldD (B). 2 μg of each purified protein were loaded on 15.4 % Tris-tricine SDS-PAGE.

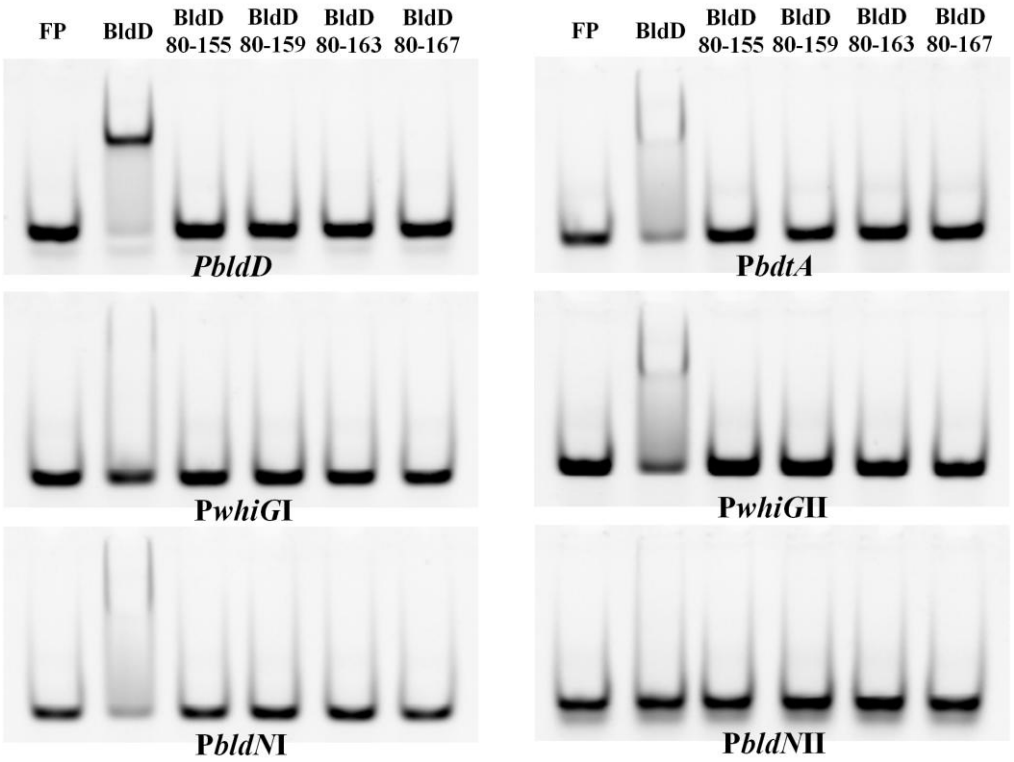


Fig. 31. Gel mobility shift assay of various truncated forms of BldD-CTD against six BldD-binding sites. 100 μM of each truncated form of BldD-CTD were mixed with each probe and loaded onto 8 % native gel containing 1X TBE. 5 μM of BldD were used as a control. Each gel was visualized with 60s exposure time in SYBR mode using LAS3000 (Fujifilm).

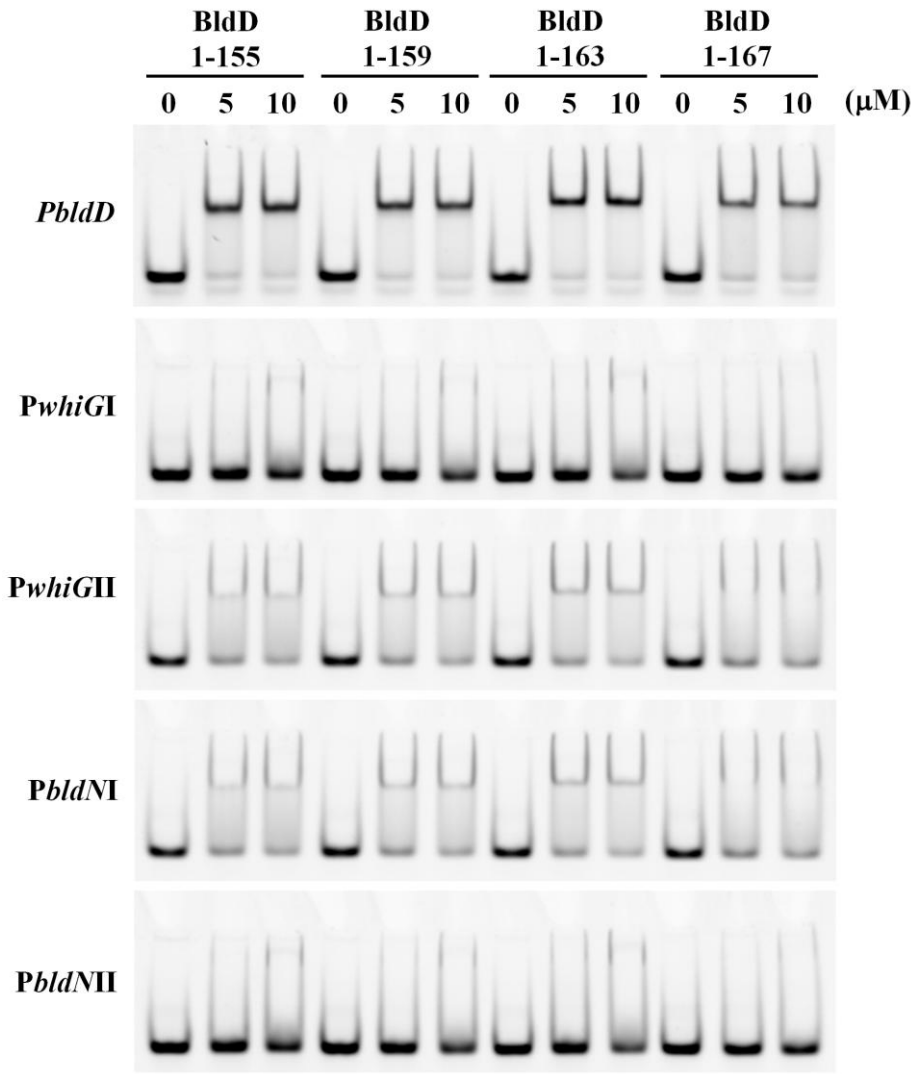


Fig. 32. Gel mobility shift assay of various truncated forms of BldD-CTD against six BldD-binding sites. 5 μM or 10 μM of each protein were mixed with each probe and loaded onto 8 % native gel containing 1X TBE. Each gel was visualized with 60s exposure time in SYBR mode using LAS3000 (Fujifilm).

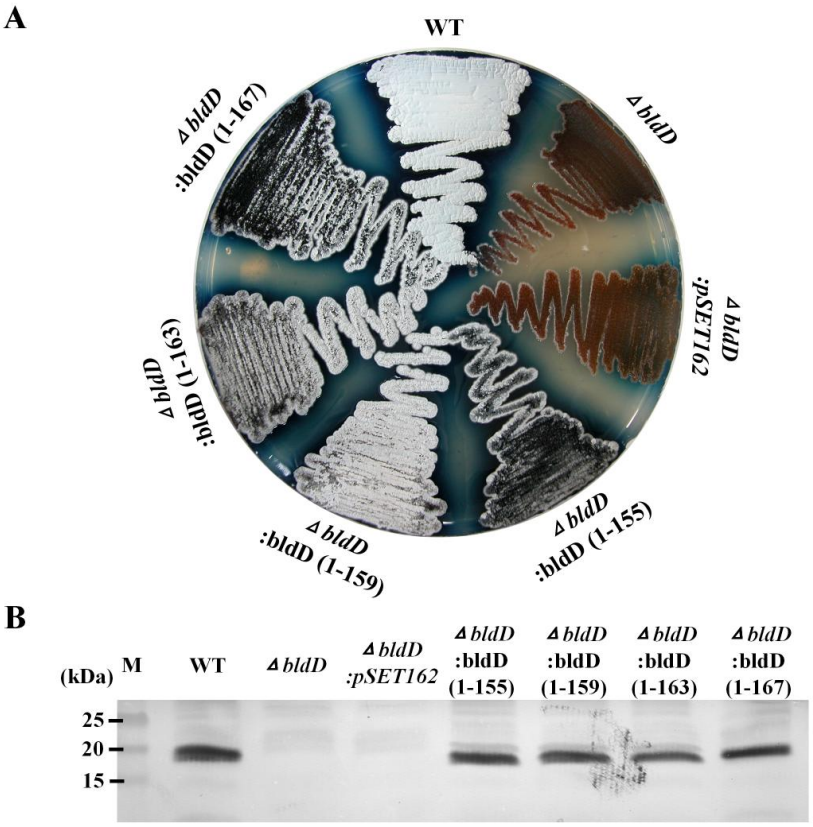


Fig. 33. *In vivo* complementation experiments of BldD truncated forms. (A) Phenotype of wild type and *ΔbldD* starains complemented by introduction of pSET162 derivatives containing wild type (WT) *bldD* and truncated *bldD*. pSET162 was used as a negative control. All strains were grown on R2YE medium for 4 days at 30 °C. (B) Immnuoblot analysis of wild type BldD and fragmented BldD. 40 μg of crude extract from each strain were loaded on each lane. All strains were harvested after grown on R2YE overlaid with cellophane disc for 1 day at 30 °C.

IV. Discussion

The solution structure of BldD-CTD has been determined and has a similar fold with winged-helix domain (WHD). The structure of the BldD-CTD contains an additional C-terminal helix, η1, which appears to be dynamic and might therefore interfere with DNA-binding of BldD-CTD. But, C-terminal truncated forms of the BldD-CTD, wherein the η1 was truncated, were also incapable of binding DNA. Generally, strongly positive-charged surface of the helix-turn-helix face is conserved in the DNA-binding WHDs and favors the interaction with the negatively charged phosphate backbone of DNA (Aravind *et al*., 2005; Harami *et al*., 2013). In contrast, the electrostatic surface potential of the BldD-CTD shows abundant negative charges rather than positive charges. Thus, anionic property of the BldD-CTD appears to be mainly responsible for its inability to bind DNA.

Although the major function of WHD is DNA-binding, WHD have variable structures and functions beyond DNA-binding. The CTD of RAP74 (the large subunit of transcription factor IIF) adopts a WHD fold and interacts with the CTD of FCP1 (RNA polymerase II C-terminal domain phosphatase) through conserved hydrophobic residues exposed from α2 and α3 (Nguyen *et al*., 2003) (Fig. 34A, 35A and 35B). Likewise, the WHD of DP2 forms

heterodimeric transcription factor with E2F4 through hydrophobic patch exposed from α1 and α3 (Zheng et al., 1999) (Fig. 34B, 35C and 35D). These suggest an alternative role for the BldD-CTD as a protein-protein interaction module. As shown in Fig. 36, the region from R105 to D128 of the BldD-CTD is strictly conserved among its homologues. In addition, ConSurf analysis (Glaser et al., 2003) also indicated highly conservation of this region (Fig. 37). Interestingly, like RAP74 and DP2, a hydrophobic surface patch is formed by the highly conserved Y106, I110, Y117, V121, L122, I124, and I135 in the BldD-CTD structure (Fig. 38). Therefore, the conserved hydrophobic surface patch on the BldD-CTD could provide a suitable interface for putative PPIs. In addition, the hydrophobic surface patch is surrounded by highly conserved charged residues, such as R105, R114, R125, D116 and D128 (Fig. 38). This could facilitate the anchoring of inbound proteins through salt-bridge formation, similar to interactions shown in the RAP74-FCP1 complex and DP2-E2F4 heterodimer (Nguyen *et al*., 2003; Zheng et al., 1999).

In conclusion, the solution structure of the BldD-CTD reveals a novel type of WHD fold that is suitable for protein-protein interactions, rather than DNA binding. Therefore, it can be suggested a possible role of the BldD-CTD as a molecular adaptor, switching transcriptional regulation through a protein-protein interaction.

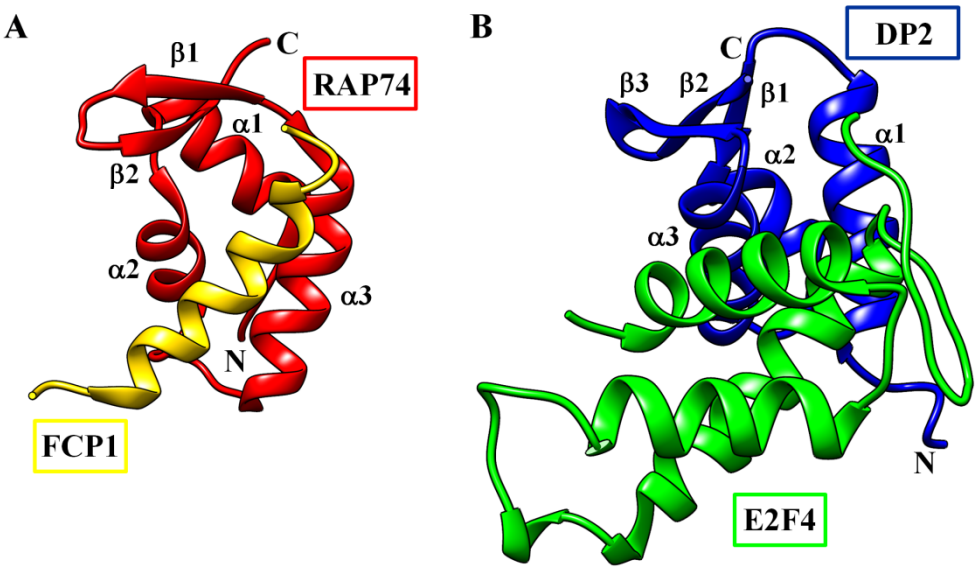


Fig. 34. Cartoon representation of RAP74-FCP1 complex and DP2-E2F4 heterodimer. (A) RAP74-FCP1 complex structure (Nguyen *et al*., 2003). RAP74-CTD and FCP1-CTD are colored red and yellow, respectively. Secondary structure elements of RAP74-CTD are labeled only. (B) DP2-E2F4 heterodimer structure (Zheng *et al*., 1999). DP2 and E2F4 are colored blue and green, respectively. Secondary structure elements of DP2 are labeled only.

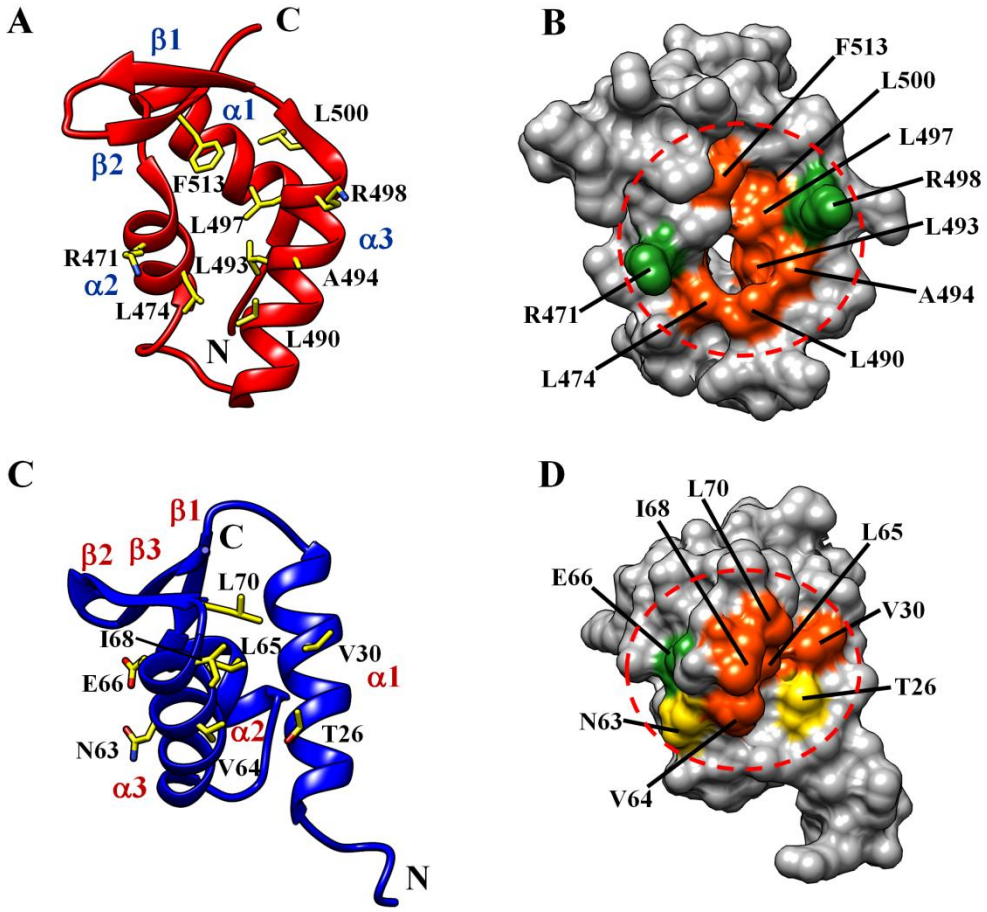


Fig. 35. Hydrophobic patch in winged-helix domain of RAP74 and DP2. Cartoon representation showing hydrophobic patch in the RAP74-CTD (A) and DP2 (C). Surface representation of RAP74-CTD (B) and DP2 (D) showing highly conserved residues. Secondary structure elements are labeled. 'N' and 'C' represent N-terminus and C-terminus, respectively. In panel (B) and (D), hydrophobic, neutral and electrically charged residues are labeled and colored orange, yellow and green, respectively.

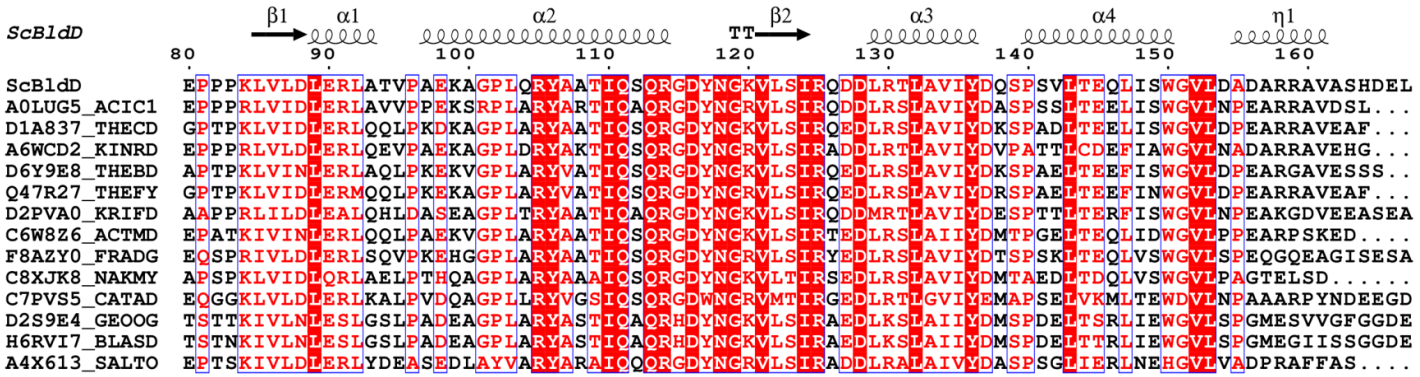

ScBldD
β1
α1
α2
TT
β2
α3
α4
η1
80
90
100
110
120
130
140
150
160
ScBldD EPPPKLVLDLERLATVPAEKAGPLQRYAATIQSQRGDYNGKVLSIRQDDLRTLAVIYDQSPSVLTEQLISWGVLDADARRAVASHDEL
A0LUG5_ACIC1 EPPPRLVIDLERLAVVPPEKSRPLARYARTIQAQRGDYNGRVLSIRQDDLRSLAVIYDASPSSLTEELISWGVLNPEARRAVDSL...
D1A837_THECD GPTPKLVIDLERLQQLPKDKAGPLARYAATIQSQRGDYNGKVLSIRQEDLRSLAVIYDKSPADLTEELISWGVLDPEARRAVEAF...
A6WCD2_KINRD EPPPRLVLDLERLQEVPAEKAGPLDRYAKTIQSQRGDYNGRVLSIRADDLRTLAVIYDVPATTLCDEFIAWGVLNADARRAVEHG...
D6Y9E8_THEBD APTPKLVINLERLAQLPKEKVGPLARYVATIQSQRGDYNGKVLSIRQEDLRSLAVIYDKSPAELTEEFISWGVLDPEARGAVESSS..
Q47R27_THEFY GPTPKLVIDLERMQQLPKEKAGPLARYVATIQSQRGDYNGRVLSIRQEDLRSLAVIYDRSPAELTEEFINWGVLDPEARRAVEAF...
D2PVA0_KRIFD AAPPRLILDLEALQHLDASEAGPLTRYAATIQAQRGDYNGKVLSIRQDDMRTLAVIYDESPTTLTERFISWGVLNPEAKGDVEEASEA
C6W8Z6_ACTMD EPATKIVINLERLQQLPAEKVGPLARYAATIQSQRGDYNGKVLSIRTEDLRSLAIIYDMTPGELTEQLIDWGVLPPEARPSKED....
F8AZY0_FRADG EQSPRIVLDLERLSQVPKEHGGPLARYAATIQSQRGDYNGRVLSIRYEDLRSLAVIYDTSPSKLTEQLVSWGVLSPEQGQEAGISESA
C8XJK8_NAKMY APSPKLVIDLQRLAELPTHQAGPLARYAAAIQSQRGDYNGKVLTIRSEDLRSLAVIYDMTAEDLTDQLVSWGVLPAGTELSD......
C7PVS5_CATAD EQGGKLVLDLERLKALPVDQAGPLLRYVGSIQSQRGDWNGRVMTIRGEDLRTLGVIYEMAPSELVKMLTEWDVLNPAAARPYNDEEGD
D2S9E4_GEOOG TSTTKIVLNLESLGSLPADEAGPLARYASTIQAQRHDYNGKVLSIRAEDLKSLAIIYDMSPDELTSRLIEWGVLSPGMESVVGFGGDE
H6RVI7_BLASD TSTNKIVLNLESLGSLPADEAGPLARYASTIQAQRHDYNGKVLSIRAEDLKSLAIIYDMSPDELTTRLIEWGVLSPGMEGIISSGGDE
A4X613_SALTO EPTSKIVLDLERLYDEASEDLAYVARYARAIQQQRGDYNGRVLSIRADDLRALAIVYDASPSGLIERLNEHGVLVADPRAFFAS....

Fig. 36. Multiple sequence alignment of the BldD-CTD with minimized redundancy. Multiple sequence alignment of the BldD-CTD and its representative homologues after removing redundancy, analyzed by ClustalW and displayed using ESPript. Secondary structure elements are illustrated over the *Streptomyces coelicolor* (Sc) BldD-CTD sequence, of which every tenth residue number is indicated. 'TT' represents a strict β-turn. Each of the other protein sequences is labeled on the left with the UniProt (http://www.uniprot.org) accession number, followed by an abbreviated name of the organism (ACIC1, *Acidothermus cellulolyticus*; THECD, *Thermomonospora curvata*; KINRD, *Kineococcus radiotolerans*; THEBD, *Thermobispora bispora*; THEFY, *Thermobifida fusca*; KRIFD, *Kribbella flavida*; ACTMD, *Actinosynnema mirum*; FRADG, *Frankia symbiont* subsp. *Datisca glomerata*; NAKMY, *Nakamurella multipartita*; CATAD, *Catenulispora acidiphila*; GEOOG, *Geodermatophilus obscurus*; BLASD, *Blastococcus saxobsidens*; SALTO, *Salinispora tropica*). The fully conserved amino acids are highlighted with red-shaded boxes, while the residues with conservative substitutions are colored red and blue-boxed.

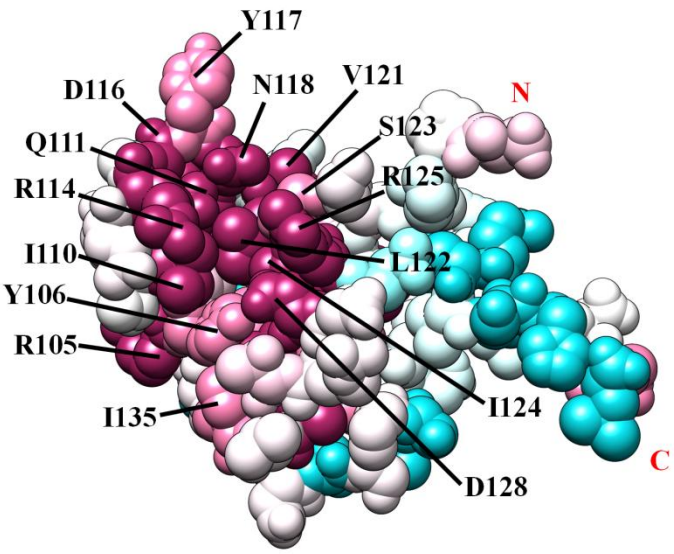


Fig. 37. Conserved surface coloring of the BldD-CTD structure. Consurf image of BldD-CTD representing the conserved residues from the multiple sequence alignment of 34 BldD-CTD homologues using ClustalW. The degree of residue conservation is colored with ranges from magenta (highly conserved) to cyan (variable). Highly conserved residues are labeled. ‘N’ and ‘C’ represent N-terminus and C-terminus, respectively.

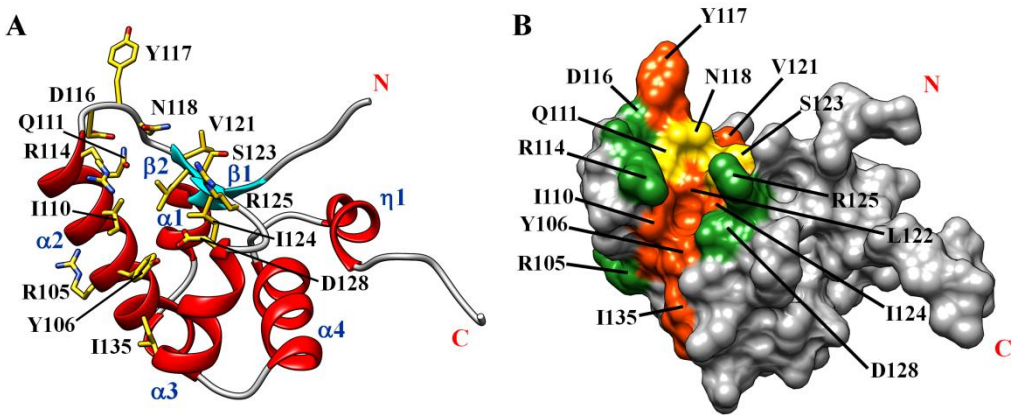


Fig. 38. Conserved hydrophobic patch of the BldD-CTD. (A) Cartoon representation showing conserved hydrophobic patch in the BldD-CTD structure. Side-chains are represented as yellow sticks with heteroatoms colored red for oxygen and blue for nitrogen. Secondary structure elements are colored red for helices (α1-α4 and η1) and cyan for β-strands (β1 and β2). (B) Surface representation of BldD-CTD showing highly conserved residues. Hydrophobic, neutral and electrically charged residues are colored orange, yellow and green, respectively. 'N' and 'C' represent N-terminus and C-terminus, respectively.

V. REFERENCES

den Hengst, C.D., Tran, N.T., Bibb, M.J., Chandra, G., Leskiw, B.K., Buttner, M.J. (2010) Genes essential for morphological development and antibiotic production in *Streptomyces coelicolor* are targets of BldD during vegetative growth. *Mol Microbiol* **78**:361-379.

Ainsa, J. A., Ryding, N. J., Hartley, N., Findlay, K. C., Bruton, C. J. & Chater, K. F. (2000). WhiA, a protein of unknown function conserved among gram-positive bacteria, is essential for sporulation in *Streptomyces coelicolor* A3(2). *J Bacteriol* **182**, 5470-5478.

Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M., Iyer, L.M. (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* **29**:231-262.

Bhattacharya, A., Tejero, R., Montelione, G.T. (2007) Evaluating protein structures determined by structural genomics consortia. *Proteins* **66**:778-795.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., Warren, G.L. (1998)

Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**:905-921.

Butala, M., Zgur-Bertok, D., Busby, S.J. (2009) The bacterial LexA transcriptional repressor. *Cell Mol Life Sci* **66**:82-93.

Champness, W. C. (1988). New loci required for *Streptomyces coelicolor* morphological and physiological differentiation. *J Bacteriol* **170**, 1168-1174.

Chater, K. F. & Merrick, M. J. (1976). Approaches to the study of differetiation in *Streptomyces coelicolor* A3(2). In *Second international symposium on the genetics of industrial microorganisms*, pp. 583-593. Edited by K. D. MacDonald. London, United Kingdom: Academic Press.

Chater, K. F. (1972). A morphological and genetic mapping study of white colony mutants of *Streptomyces coelicolor*. *J Gen Microbiol* **72**, 9-28.

Chater, K. F. (1993). Genetics of differentiation in Streptomyces. *Annu Rev Microbiol* **47**, 685-713.

Chater, K. F., Bruton, C. J., Plaskitt, K. A., Buttner, M. J., Mendez, C. and Helmann, J. D. (1989) The developmental fate of *S. coelicolor* hyphae depends upon a gene product homologous with the motility factor of *B. subtilis*. *Cell* **59 :** 133-143.

Davis, N. K. & Chater, K. F. (1992). The *Streptomyces coelicolor whiB* gene

encodes a small transcription factor-like protein dispensable for growth but essential for sporulation. *Mol Gen Genet* **232**, 351-358.

Delaglio, F., Grzesiek, S., Vuister, GW., Zhu, G., Pfeifer, J., Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**:277-293.

Elliot, M. A. & Leskiw, B. K. (1999). The BldD protein from *Streptomyces coelicolor* is a DNA-binding protein. *J Bacteriol* **181**, 6832-6835.

Elliot, M., Damji, F., Passantino, R., Chater, K. & Leskiw, B. (1998). The *bldD* gene of *Streptomyces coelicolor* A3(2): a regulatory gene involved in morphogenesis and antibiotic production. *J Bacteriol* **180**, 1549-1555.

Elliot, M. A., Bibb, M. J., Buttner, M. J. & Leskiw, B. K. (2001). BldD is a direct regulator of key developmental genes in *Streptomyces coelicolor* A3(2). *Mol Microbiol* **40**, 257-269.

Elliot, M.A., Locke, T.R., Galibois, C.M., Leskiw, B.K.(2003) BldD from *Streptomyces coelicolor* is a non-essential global regulator that binds its own promoter as a dimer. *FEMS Microbiol Lett* **225**:35-40.

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., Kay, L.E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry*

33:5984-6003.

Fernandez-Moreno, M. A., Caballero, J. L., Hopwood, D. A. & Malpartida, F. (1991). The act cluster contains regulatory and antibiotic export genes, direct targets for translational control by the *bldA* tRNA gene of Streptomyces. *Cell* **66**, 769-780.

Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., Ben-Tal, N. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **19**:163-164.

Güntert, P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**:353-378.

Güntert, P., Mumenthaler, C., Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* **273**:283-298.

Harami, G.M., Gyimesi, M., Kovács, M. (2013) From keys to bulldozers: expanding roles for winged helix domains in nucleic-acid-binding proteins. *Trends Biochem Sci* **38**:364-371.

Herrmann, T., Güntert, P., Wüthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol*

319:209-227.

Holm, L., Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**:W545-549.

Hopwood, D. A., Wildermuth, H. & Palmer, H. M. (1970). Mutants of *Streptomyces coelicolor* defective in sporulation. *J Gen Microbiol* **61**, 397-408.

Ikura, M., Kay, L.E., Bax, A. (1991) Improved three-dimensional 1H-13C-1H correlation spectroscopy of a 13C-labeled protein using constant-time evolution. *J Biomol NMR* **1**:299-304.

Johnson, B. A. (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* **278**:313-352.

Kay, L.E., Ikura, M., Tschudin, R., Bax, A. (1990) Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. *J Magn Reson* **213**:423-441.

Kelemen, G. H. & Buttner, M. J. (1998). Initiation of aerial mycelium formation in Streptomyces. *Curr Opin Microbiol* **1**, 656-662.

Kelemen, G. H., Brown, G. L., Kormanec, J., Potuckova, L., Chater, K. F. & Buttner, M. J. (1996). The positions of the sigma-factor genes, *whiG* and *sigF*, in the hierarchy controlling the development of spore chains in the aerial hyphae of *Streptomyces coelicolor* A3(2). *Mol Microbiol* **21**,

593-603.

Kelemen, G. H., Viollier, P. H., Tenor, J., Marri, L., Buttner, M. J. & Thompson, C. J. (2001). A connection between stress and development in the multicellular prokaryote *Streptomyces coelicolor* A3(2). *Mol Microbiol* **40**, 804-814.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., Hopwood, D.A. (2000) Practical *Streptomyces* Genetics. *The John Innes Foundation Norwich.*

Kim, I.K., Lee, C.J., Kim, M.K., Kim, J.M., Kim, J.H., Yim, H.S., Cha, S.S., Kang, S.O. (2006) Crystal structure of the DNA-binding domain of BldD, a central regulator of aerial mycelium formation in *Streptomyces coelicolor* A3(2). *Mol Microbiol* **60**:1179-1193.

Lawlor, E. J., Baylis, H. A. & Chater, K. F. (1987). Pleiotropic morphological and antibiotic deficiencies result from mutations in a gene encoding a tRNA-like product in *Streptomyces coelicolor* A3(2). *Genes Dev* **1**, 1305-1310.

Lee, C.J., Won, H.S., Kim, J.M., Lee, B.J., Kang, S.O. (2007) Molecular domain organization of BldD, an essential transcriptional regulator for developmental process of *Streptomyces coelicolor* A3(2). *Proteins* **68**:344-352.

Leskiw, B. K., Lawlor, E. J., Fernandez-Abalos, J. M. & Chater, K. F.

(1991). TTA codons in some genes prevent their expression in a class of developmental, antibiotic-negative, Streptomyces mutants. *Proc Natl Acad Sci U S A* **88**, 2461-2465.

Linge, J.P., Williams, M.A., Spronk, C.A., Bonvin, A.M., Nilges, M. (2003) Refinement of protein structures in explicit solvent. *Proteins* **50**:496-506.

Losick, R. & Shapiro, L. (1993). Checkpoints that couple gene expression to morphogenesis. *Science* **262**, 1227-1228.

Markley, J.L., Bax, A., Arata, Y., Hilbers, C.W., Kaptein, R., Sykes, B.D., Wright, P.E., Wüthrich, K. (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. *J Mol Biol* **280**:933-952.

Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M., Clore, G.M. (1989) Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry* **28**:6150-6156.

Merrick, M. J. (1976). A morphological and genetic mapping study of bald colony mutants of *Streptomyces coelicolor*. *J Gen Microbiol* **96**, 299-315.

Molle, V. & Buttner, M. J. (2000). Different alleles of the response regulator gene *bldM* arrest *Streptomyces coelicolor* development at distinct stages.

Mol Microbiol **36**, 1265-1278.

Molle, V., Palframan, W. J., Findlay, K. C. & Buttner, M. J. (2000). WhiD and WhiB, homologous proteins required for different stages of sporulation in *Streptomyces coelicolor* A3(2). *J Bacteriol* **182**, 1286-1295.

Nederveen, A.J., Doreleijers, J.F., Vranken, W., Miller, Z., Spronk, C.A., Nabuurs, S.B., Güntert, P., Livny, M., Markley, J.L., Nilges, M., Ulrich, E.L., Kaptein, R., Bonvin, A.M. (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* **59**:662-672.

Nepravishta, R,, Polizio, F., Paci, M., Melino, S. (2012) A metal-binding site in the RTN1-C protein: new perspectives on the physiological role of a neuronal protein. *Metallomics* 4:480-487.

Nguyen, B.D., Abbott, K.L., Potempa, K., Kobor, M.S., Archambault, J., Greenblatt, J., Legault, P., Omichinski, J.G. (2003) NMR structure of a complex containing the TFIIF subunit RAP74 and the RNA polymerase II carboxyl-terminal domain phosphatase FCP1. *Proc Natl Acad Sci USA* **100**:5688-5693.

Nodwell, J. R., McGovern, K. & Losick, R. (1996). An oligopeptide permease responsible for the import of an extracellular signal governing aerial mycelium formation in *Streptomyces coelicolor*. *Mol Microbiol* **22**, 881-

893.

Nodwell, J. R., Yang, M., Kuo, D. & Losick, R. (1999). Extracellular complementation and the identification of additional genes involved in aerial mycelium formation in *Streptomyces coelicolor*. *Genetics* **151**, 569-584.

Olejniczak, E.T., Xu, R.X., Fesik, S.W. (1992) A 4D HCCH-TOCSY experiment for assigning the side chain 1H and 13C resonances of proteins. *J Biomol NMR* **2**:655-659.

Pope, M. K., Green, B. & Westpheling, J. (1998). The *bldB* gene encodes a small protein required for morphogenesis, antibiotic production, and catabolite control in *Streptomyces coelicolor*. *J Bacteriol* **180**, 1556-1562.

Ryding, N. J., Kelemen, G. H., Whatling, C. A., Flardh, K., Buttner, M. J. & Chater, K. F. (1998). A developmentally regulated gene encoding a repressor-like protein is essential for sporulation in *Streptomyces coelicolor* A3(2). *Mol Microbiol* **29**, 343-357.

Ryding, N. J., Bibb, M. J., Molle, V., Findlay, K. C., Chater, K. F. & Buttner, M. J. (1999). New sporulation loci in *Streptomyces coelicolor* A3(2). *J Bacteriol* **181**, 5419-5425.

Shen, Y., Delaglio, F., Cornilescu, G., Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR

chemical shifts. *J Biomol NMR* **44**:213-223.

Skjaerbaek, N., Nielsen, K.J., Lewis, R.J., Alewood, P., Craik, D.J. (1997) Determination of the solution structures of conantokin-G and conantokin-T by CD and NMR spectroscopy. *J Biol Chem* **272**:2291-2299

Stoll, K.E., Draper, W.E., Kliegman, J.I., Golynskiy, M.V., Brew-Appiah, R.A., Phillips, R.K., Brown, H.K., Breyer, W.A., Jakubovics, N.S., Jenkinson, H.F., Brennan, R.G., Cohen, S.M., Glasfeld, A. (2009) Characterization and structure of the manganese-responsive transcriptional regulator ScaR. *Biochemistry* **48**:10308-10320.

Tan, H., Yang, H., Tian, Y., Wu, W., Whatling, C. A., Chamberlin, L. C., Buttner, M. J., Nodwell, J. & Chater, K. F. (1998). The *Streptomyces coelicolor* sporulation-specific σ^{WhiG} form of RNA polymerase transcribes a gene encoding a ProX-like protein that is dispensable for sporulation. *Gene* **212**, 137-146.

Tillotson, R. D., Wosten, H. A., Richter, M. & Willey, J. M. (1998). A surface active protein involved in aerial hyphae formation in the filamentous fungus *Schizophillum commune* restores the capacity of a bald mutant of the filamentous bacterium *Streptomyces coelicolor* to erect aerial structures. *Mol Microbiol* **30**, 595-602.

Willey, J., Santamaria, R., Guijarro, J., Geistlich, M. & Losick, R. (1991).

Extracellular complementation of a developmental mutation implicates a small sporulation protein in aerial mycelium formation by *S. coelicolor*. *Cell* **65**, 641-650.

Willey, J., Schwedock, J. & Losick, R. (1993). Multiple extracellular signals govern the production of a morphogenetic protein involved in aerial mycelium formation by *Streptomyces coelicolor*. *Genes Dev* **7**, 895-903.

Wishart, D.S., Sykes, B.D. (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR* **4**:171-180.

Wishart, D.S., Sykes, B.D., Richards, F.M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**:1647-1651.

Wüthrich, K. (1986) NMR of Proteins and Nucleic Acids. *John Wiley & Sons, Inc.*

Zheng, N., Fraenkel, E., Pabo, C.O., Pavletich, N.P. (1999) Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. *Genes Dev* **13**:666-674.

국문초록

BldD 단백질은 토양세균인 방선균 내에 존재하는 단백질로 핵산과 결합하는 기능을 가지며, 167개의 아미노산 서열로 구성되어 있다. 방선균 중의 하나인 *Streptomyces coelicolor*는 외부 환경이나 영양 요소의 이용여부에 따라 포자를 형성한다. 이러한 포자 형성은 우선적으로 기균사 형성의 선행이 필수적이다. 이러한 기균사의 형성에 BldD 단백질의 역할이 상당히 중요하다는 많은 보고가 있어 왔고, 실제로 이 BldD 단백질은 분화와 관련된 *bldN* 이나 *whiG* 과 같은 유전자들을 세포 생장 초기에 억제하는 것으로 밝혀졌다. BldD 단백질은 두 개의 독립된 기능체로 이루어져 있으며, 그 중 아미노 말단에 가까이 위치하는 기능체는 BldD 단백질이 핵산과 결합함에 있어 주된 역할을 하는 것이 본 실험실의 연구를 통해 밝혀졌다. 하지만, 카르복시 말단에 가까이 위치하는 기능체는 BldD 단백질이 기능하는데 있어 어떠한 역할을 하고 있는지에 관한 연구 결과는 현재까지 없는 실정이다. 또한 본 실험실의 연구에 의해 BldD 단백질의 이량체 형성에도 아미노 말단 지역의 기능체가 작용하는 것으로 밝혀졌다. 여러 가지 생리학적 또는 생화학적 방법론을 통해 BldD 단백질의 카르복시 말단 지역 기능체에 대한 다양한 연구가 본 실험실에서 시도

되었으나 그 기능을 정확히 동정할 수 없었다. 따라서, 현재까지 시도되었던 생화학적 방법론이 아닌 생물물리학적 접근이 필요한 것으로 생각되었고, 그 중 용액 상태에서의 단백질의 삼차 구조 연구가 가능한 자기공명장치를 이용하여 BldD 단백질의 카르복실 말단 지역 기능체의 삼차 구조를 밝혀내었다. 놀랍게도 그 삼차 구조는 기존에 밝혀졌던 단백질의 구조와는 상이한 새로운 구조인 것으로 밝혀졌으며, 예상과 달리 핵산과 결합할 수 있는 다른 단백질과 높지는 않으나 상당한 정도의 유사성을 보였다. 하지만, 구조적인 유사성과 달리 BldD 단백질의 카르복시 말단 지역 기능체는 일반적인 핵산 결합 단백질과 대조적으로 음전하의 표면을 가지고 있으며, 이 음전하의 분포로 인해 인산에 의해 강한 음전하를 가지는 핵산과 결합할 수 있는 가능성은 매우 적은 것으로 여겨지며, 이것은 젤 이동성 변이 실험을 통해 확인할 수 있었다. 또한 이 단백질은 잘 보존되었고, 외부로 노출되어 있는 소수성 아미노산 집합체를 가지고 있는 것이 구조를 통해 분석되었고, 이러한 소수성 아미노산 집합체의 존재로 볼 때 BldD 단백질의 카르복시 말단 지역 기능체는 다른 단백질과의 상호작용을 통해 BldD 단백질의 전체 기능을 조절하는 것으로 예상된다.

감사의 글

이제는 기억도 흐릿해질 정도의 많은 시간이 걸려 한참이나 부족하지만 한 편의 논문으로 정리하여 박사학위과정을 마무리하게 되었습니다. 과학자로서의 길을 나아감에 있어 남들보다 조금은 늦게 첫 번째 이정표를 지나며, 한 번 뒤돌아보니 많은 아쉬움도 남지만, 무엇보다 결코 혼자서는 이 길을 걸어오지 못했다는 생각이 강하게 듭니다.

먼저, 많이 부족하고 또한 느리디 느린 저를 끝없는 포용과 관심으로 이끌어 주시고, 진정한 과학의 길에 대한 독보적인 철학이 묻어나는 끊임없는 교육으로 제가 앞으로 걷게 될 길의 든든한 반석을 마련해 주신 은사님 강사욱 선생님께 깊은 감사를 올립니다. 실험을 함에 있어 큰 흐름과 그것을 이루는 작은 물줄기 모두를 두루 살펴 주시며 제게 해주셨던 그 모든 말씀들은 제 삶에 있어 큰 밑거름이 될 것이라 확신합니다. 바쁘신 중에도 제 논문을 면밀하게 살펴주시고, 그에 따른 여러 가지 조언들을 아끼지 않고 해주신 석영재 교수님, 허원기 교수님 그리고 최희정 교수님께도 깊은 감사를 드립니다. 그리고, NMR에 대해 이해조차 제대로 못했던 저에게 많은 시간을 할애하여 설명해주시고, NMR 실험과 분석 전반에 걸쳐 도움 주시는 것을 주저하지 않으시며, 논문 작성에 있어서도 지도와 조언을 끊임없이 해주신 원형식 교수님께 깊은 감사를 올립니다. 또한, 자포자기하며 나락으로 빠져버리려던 저에게 끝없는 조언으로 희망을 갖게 해주셨던 임형순 선생님, 정신 누님께 깊이 감사 드립니다.

생물물리학 연구실이라는 같은 공간에서 생활하며 저에게 알게 모르게 많은 도움을 주었던 선배님들과 후배님들께도 감사를 전하고 싶습니다. 선배님들께서 하시는 말씀은 잘 듣지도 않고 하고픈 데로 하고 살았었던 철없던 저를 넓은 마음으로 이해해 주시고, 많은 조언을 해주셨던 모든 선배님들께 깊이 감사 드립니다. 또한, 그다지 좋은 선배의 모습을 보여주지 못하고 변변한 도움조차 주지 못하면서, 부담스럽고 까다로운 선배의 모습만을 보여 주었으나, 그런 선배를 이해해주고 받아주었던 모든 후배님들께도 깊은 감사를 전합니다.

어떠한 말로도 그 은혜를 다 갚을 길이 없겠지만, 이 길을 걸어오기까지 무엇 하나 제대로 된 모습을 보여드리지 못해도, 항상 저를 믿고 묵묵히 지켜봐 주시며, 뒷바라지 하느라 끝없이 희생하신 아버님, 어머님께 깊이 감사를 올립니다. 못난 사위임에도 불구하고 항상 염려해 주시고, 아내보다 저를 더 많이 챙겨주시며, 또 한 번 부모님의 따뜻함을 느낄 수 있게 해주신 장인어른과 장모님께도 깊은 감사를 드립니다.

아빠가 무슨 일을 하는지 잘 모르겠지만, 논문 준비하는 중에 태어나 고생하면서도, 튼튼하고 씩씩하게 자라며 방긋방긋 웃어주는 아들 준우에게 아빠가 많이 고맙고, 많이 사랑한다고 이야기해주고 싶습니다. 마지막으로 지금의 제가 있고, 앞으로의 제가 있기 위한 반석이자 도달점인 사랑하는 아내 지은에게 말로는 다 표현 못할 고마움을 전합니다.